

## (19) 대한민국특허청(KR)

## (12) 공개특허공보(A)

(51) Int. Cl. <sup>7</sup> G11B 20/10	(11) 공개번호 (43) 공개일자	특2003-0059085 2003년07월07일
(21) 출원번호	10-2003-7001489	
(22) 출원일자	2003년01월30일	
번역문 제출일자	2003년01월30일	
(86) 국제출원번호	PCT/EP2001/08709	
(86) 국제출원일자	2001년07월26일	
(87) 국제공개번호	WO 2002/11123	
(87) 국제공개일자	2002년02월07일	
(81) 지정국	<p>국내특허: 알바니아, 아르메니아, 오스트리아, 오스트레일리아, 아제르바이잔, 보스니아-헤르체고비나, 바베이도스, 불가리아, 브라질, 벨라루스, 캐나다, 스위스, 중국, 쿠바, 체코, 독일, 덴마크, 에스토니아, 스페인, 핀란드, 영국, 그루지야, 헝가리, 이스라엘, 아이슬란드, 일본, 케냐, 키르기즈, 북한, 대한민국, 카자흐스탄, 세인트루시아, 스리랑카, 라이베리아, 레소토, 리투아니아, 룩셈부르크, 라트비아, 몰도바, 마다가스카르, 마케도니아, 몽고, 말라위, 멕시코, 노르웨이, 뉴질랜드, 슬로베니아, 슬로바키아, 타지키스탄, 투르크메니스탄, 터어키, 트리니다드토바고, 우크라이나, 우간다, 미국, 우즈베키스탄, 베트남, 폴란드, 포르투갈, 루마니아, 러시아, 수단, 스웨덴, 싱가포르, 아랍에미리트, 안티구아바부다, 코스타리카, 도미니카연방, 알제리, 모로코, 탄자니아, 남아프리카, 벨리즈, 모잠비크, 콜롬비아, 그레나다, 가나, 감비아, 크로아티아, 인도네시아, 인도, 유고슬라비아, 짐바브웨</p> <p>AP ARIPO특허: 케냐, 레소토, 말라위, 수단, 스와질랜드, 우간다, 시에라리온, 가나, 감비아, 짐바브웨, 탄자니아, 모잠비크</p> <p>EA 유라시아특허: 아르메니아, 아제르바이잔, 벨라루스, 키르기즈, 카자흐스탄, 몰도바, 러시아, 타지키스탄, 투르크메니스탄</p> <p>EP 유럽특허: 오스트리아, 벨기에, 스위스, 독일, 덴마크, 스페인, 프랑스, 영국, 그리스, 아일랜드, 이탈리아, 룩셈부르크, 모나코, 네덜란드, 포르투갈, 스웨덴, 핀란드, 사이프러스, 터어키</p> <p>OA OAPI특허: 부르키나파소, 베닝, 중앙아프리카, 콩고, 코트디부아르, 카메룬, 가봉, 기네, 말리, 모리타니, 니제르, 세네갈, 차드, 토고, 기네비쏘, 적도기네</p>	
(30) 우선권주장	60/222,023 2000년07월31일 미국(US) 09/839,476 2001년04월20일 미국(US)	
(71) 출원인	<p>샤잠 엔터테인먼트 리미티드</p> <p>영국</p> <p>영국 런던 스위트 22 워더 스트리트 189</p>	
(72) 발명자	<p>왕에이버리리춘</p> <p>미국</p> <p>미국캘리포니아주팔로알토로스로드2915</p> <p>스미스줄리어스오.3세</p> <p>미국</p> <p>미국캘리포니아주팔로알토밀러애버뉴4360</p>	
(74) 대리인	<p>김명신</p> <p>이기성</p> <p>김중호</p> <p>박장규</p>	
(77) 심사청구	없음	
(54) 출원명	오디오 데이터베이스에서의 검색 방법	

## 요약

오디오 샘플을 인식하기 위한 방법은 많은 세트의 원래 레코딩을 인덱싱하는 데이터베이스로부터 상기 오디오 샘플에 가장 가깝게 매칭되는 오디오 파일을 찾는다. 각각의 인덱싱된 오디오 파일이 데이터베이스 색인 내에 한 세트의 랜드마크 타임포인트 및 관련된 핑거프린트에 의해 표현된다. 랜드마크는 파일 내의 재생 가능한 위치에서 발생되고, 핑거프린트는 상기 랜드마크 타임포인트 또는 그 근방에서의 신호의 특징을 나타낸다. 인식을 실행하기 위하여, 알려지지 않은 샘플에 대해 랜드마크와 핑거프린트가 계산되어, 상기 데이터베이스로부터 매칭 핑거프린트를 검색하기 위해 사용된다. 매칭 핑거프린트를 포함하는 각각의 파일에 대하여, 랜드마크는 동일한

핑거프린트가 계산된 샘플의 랜드마크와 비교된다. 아주 많은 수의 대응하는 랜드마크가 선행적으로 관련되면, 즉, 샘플 및 검색된 파일의 동일한 핑거프린트가 동일한 시간 변화를 갖는다면, 파일은 샘플과 동일한 것으로 간주된다. 임의의 타입의 사운드 또는 음악에 대하여 상기 방법이 사용될 수 있고, 상기 방법은 배경 잡음, 압축 인공물 또는 전송 드롭아웃과 같은 선행 및 비선형 왜곡을 받는 오디오 신호에 특히 효과적이다. 샘플은 데이터베이스 내의 엔트리 수의 로그값에 비례하는 시간에서 식별될 수 있다. 충분한 계산 가능한 전력이 제공되는 경우, 사운드가 샘플링될 때 거의 실시간으로 인식이 실행될 수 있다.

## 대표도

### 도1

## 명세서

### 기술분야

본 발명은 일반적으로 콘텐츠 기반 정보 검색에 관한 것이다. 보다 구체적으로는 상당히 왜곡되거나 높은 잡음 레벨을 담고 있는, 사운드 또는 음악을 포함하는 오디오 신호의 인식에 관한 것이다.

### 배경기술

음악 또는 다양한 공급원으로부터 발생된 다른 오디오 신호의 자동 인식에 대한 필요성이 증가하고 있다. 예를 들어, 저작권이 있는 작품의 소유자나 광고주들은 그들의 작품의 데이터를 방송 주파수 통하여 얻는데 흥미를 갖는다. 음악 트래킹 서비스는 광범위한 시장에 주요한 방송국의 플레이리스트(playlist)를 제공한다. 소비자들은 라디오를 통해서 노래와 광고 방송을 확인하고 싶어하므로, 그들은 새롭고 흥미로운 음악 또는 다른 제품 및 서비스를 구매할 수 있게 된다. 인간이 실행할 때, 어떤 종류의 연속적인 또는 요구에 의한 음악의 인식은 효과가 없고 노동 집약적이 된다. 따라서 음악 또는 사운드를 자동으로 인식하는 방법은 소비자, 예술가 및 다양한 산업에 상당한 이익을 제공할 수 있을 것이다. 음악 배급 패러다임이 상점 구매 방식에서 인터넷을 통한 다운로드 방식으로 이동함에 따라, 컴퓨터에 의해 실행되는 음악 인식을 인터넷 구매 및 다른 인터넷 기반 서비스와 직접 연결하는 것은 상당히 실행 가능성이 있다.

종래, 라디오를 통해 들리는 노래의 인식은 노래가 연주된 횟수와 방송국을 상기 방송국 또는 제 3의 공급원에 의해 제공된 플레이리스트와 매칭함으로써 실행되었다. 이 방법은 정보가 이용될 수 있는 방송국에만 본질적으로 제한된다. 다른 방법들은 방송 신호 내에 삽입된 들을 수 없는 코드에 의존한다. 상기 삽입된 신호는 상기 방송 신호에 대한 식별 정보를 추출하기 위하여 수신기에서 디코딩된다. 이 방법의 단점은 신호를 식별하기 위해 특별한 디코딩 장치가 필요하고, 삽입 코드를 갖는 신호들만이 식별될 수 있다는 것이다.

임의의 대규모 오디오 인식은 여러 종류의 콘텐츠 기반 오디오 검색을 필요로 하는데, 이것에서는 식별되지 않은 방송 신호는 알려진 신호의 데이터베이스와 비교되어 유사 또는 동일한 데이터베이스 신호를 식별한다. 콘텐츠 기반 오디오 검색은 오디오 파일과 관련되거나 오디오 파일을 둘러싸는 메타데이터 텍스트만이 검색되는 웹 검색 엔진에 의한 현존의 오디오 검색과는 다르다는 것에 주목하자. 또한, 음성 인식이 음성 신호를, 공지된 기술을 사용하여 검색 또는 인덱싱될 수 있는 텍스트로 변환하는데 유용한 반면, 이것은 음악과 사운드를 담고 있는 많은 오디오 신호에 적용될 수 없다는 것에 주목하자. 여러 가지 점에서, 오디오 정보 검색은 검색 엔진에 의해 제공된 텍스트 기반 정보 검색과 동일하다. 다른 점에서, 그러나 오디오 인식은 유사하지 않은데, 오디오 신호는 검색 및 인덱싱을 위한 식별자를 제공하는 단어와 같은 쉽게 식별할 수 있는 엔티티가 없다. 이와 같이, 현재의 오디오 검색 방법은 신호의 다양한 성질 또는 특성을 나타내는 연산된 지각 특징에 의해 오디오 신호를 인덱싱한다.

콘텐츠 기반 오디오 검색은 통상적으로 다수의 표현 특징을 얻기 위해 조회 신호(query signal)를 분석하고, 유사성 정도를 얻어진 특징에 적용하여 상기 조회 신호와 가장 유사한 데이터베이스 파일을 찾아냄으로써 실행된다. 수신된 객체의 유사성은 선택된 지각 특징의 닮은 정도이다. 당 분야에서 다수의 콘텐츠 기반 검색 방법이 이용될 수 있다. 예를 들어, Kenyon에 의해 출원된 미국 특허 번호 5,210,820호에는 수신된 신호가 처리되고 샘플링되어 각각의 샘플링 포인트에서 신호 값이 얻어지는 신호 인식 방법이 개시되어 있다. 그 다음, 샘플링된 값의 통계적 모멘트가 계산되어, 동일한 신호를 검색하기 위하여 저장된 신호의 식별자들과 비교될 수 있는 특성 벡터가 생성된다. Kenyon 등에 의해 출원된 미국 특허 번호 4,450,531호와 4,843,562호에는 식별되지 않은 신호들과 저장된 기존 신호들 사이의 교차 상호 관계가 계산되는 유사한 방송 정보 분류 방법이 개시되어 있다.

음향의 유사성에 의해 오디오 다큐먼트를 검색하는 시스템이 C.-C.J.Kuo 등에 의해 J.T.Foote의 Multimedia Storage and Archiving Systems II, Proc. of SPIE, 3329권, 138-147 페이지(1997)에 "Content-Based Retrieval of Music and Audio"의 제목으로 개시되어 있다. 특성 벡터들은 각각의 오디오 파일을 멜-스케일 셉스트럴 계수(mel-scaled cepstral coefficient, MSCC)의 파라미터로 나타냄으로써 계산되고, 양자화 트리가 상기 파라미터화 데이터로부터 성장된다. 조회를 실행하기 위해, 특성 벡터를 얻기 위해 알려지지 않은 신호가 파라미터화되어 상기 트리의 나뭇잎 노드로 정렬된다. 각각의 나뭇잎 노드에 대해 막대그래프가 수집되어, 상기 알려지지 않은 신호를 나타내는 N 차원의 벡터가 생성된다. 이와 같은 두 개의 벡터 간의 거리는 2개의 사운드 파일 사이의 유사성을 나타낸다. 이 방법에서, 관리된 양자화 계획은 인간에 의해 트레이닝 데이터가 할당되는 클래스에 기초하여 중요하지 않은 변수를 무시하면서 특징적인 오디오 특성을 인식한다. 분류 시스템에 따라서, 다른 음향 특성이 중요한 것으로 선택된다. 따라서 이 방법은 음악을 인식하는 것보다는 노래 간의 유사성을 확인하고 음악을 클래스로 분류하는 것에 적절하다.

오디오 정보의 세그먼트, 콘텐츠 기반 분석, 기억 및 검색을 위한 방법이 Blum 등에 의해 출원된 미국 특허 번호 5,918,222호에 개시되어 있다. 이 방법에서, 음의 크기, 베이스, 피치, 휘도, 대역폭 및 MFCC(Mel-frequency cepstral coefficient)와 같은 다수의 음향 특성이 각각의 파일의 주기적인 간격에서 측정된다. 특성 벡터를 형성하기 위해 상기 특징의 통계적인 측정이 이루어지고 결합된다. 데이터베이스 내의 오디오 데이터 파일은 식별되지 않은 파일의 특성 벡터에 대한 이들 특성 벡터의 유사성에 기초하여 검색된다.

상기의 모든 종래 기술의 오디오 인식 방법의 중요한 문제점은 인식될 신호들이 예를 들어 배경 잡음, 전송 에러와 드롭아웃, 간섭, 대역 제한 필터링, 양자화, 시간 왜곡 및 음질의 디지털 압축에 의해 발생된 선형 및 비선형 왜곡을 받을 때 이들 방법이 실패하는 경향이 있다는 것이다. 종래 기술의 방법에서, 왜곡된 사운드 샘플이 처리되어 음향적인 특성이 얻어지는 경우, 원래의 레코딩에 대해 얻어진 일부분의 특징만이 발견된다. 그러므로 그 결과의 특성 벡터가 원래의 레코딩의 특성 벡터와 상당히 유사하지 않아서, 정확한 인식이 실행될 수 없을 것이다. 높은 잡음 및 왜곡의 조건 하에서 충분히 실행되는 사운드 인식 시스템에 대한 필요성이 남아있다.

종래 기술의 방법이 갖는 또 다른 문제는 그들이 계산상 철저하고, 충분히 스케일링되지 않는다는 것이다. 따라서 큰 데이터베이스를 갖는 종래 기술 방법을 사용하여 실시간 인식을 하는 것이 불가능하다. 이와 같은 시스템에서, 수백 또는 수천 개 이상의 레코딩의 데이터베이스를 갖는 것은 실행 불가능하다. 종래 기술의 방법에서 검색 시간은 데이터베이스의 크기에 따라 선형적으로 증가하여, 수백만의 사운드 레코딩에 대한 스케일링이 경제적으로 실행 불가능하게 된다. 상기 Kenyon의 방법 또한 큰 전문화된 디지털 신호 처리 하드웨어의 बैं크를 필요로 한다.

현존하는 사용되는 방법들은 종종 인식을 실행할 수 있는 입력 샘플에 대한 엄격한 요구 조건을 갖는다. 예를 들어, 이들은 전체 노래 또는 적어도 30초의 노래가 샘플링될 것을 요구하거나, 시작부터 노래가 샘플링될 것을 요구한다. 이들은 또한 하나의 스트림에 서로 혼합된 다수의 노래를 인식하는데 어려움을 갖는다. 이들의 모든 단점들 때문에 종래 기술의 방법이 많은 실질적인 응용의 사용에서 실행되지 못한다.

#### 목적 및 이점

따라서 본 발명의 주요한 목적은 높은 레벨의 잡음 및 왜곡을 받는 오디오 신호를 인식하는 방법을 제공하는 것이다.

본 발명의 다른 목적은 수초의 식별된 신호에만 기초하여 실시간으로 실행될 수 있는 인식 방법을 제공하는 것이다.

본 발명의 또 다른 목적은 단지 시작 시가 아닌 사운드 내의 거의 모든 곳으로부터의 샘플에 기초하여 사운드를 인식할 수 있는 인식 방법을 제공하는 것이다.

본 발명의 또 다른 목적은 사운드 샘플이 특정 방송국 또는 플레이리스트와 상관되거나 코딩될 것을 요구하지 않는 인식 방법을 제공하는 것이다.

본 발명의 또 다른 목적은 하나의 스트림에 함께 혼합된 다수의 사운드 레코딩의 각각을 인식할 수 있는 인식 방법을 제공하는 것이다.

본 발명의 또 다른 목적은 알려지지 않은 사운드가 실질적으로 알려진 방법에 의해 임의의 환경으로부터 시스템에 제공될 수 있는 사운드 인식 시스템을 제공하는 것이다.

#### 발명의 상세한 설명

이들 목적과 이점은 다수의 알려진 미디어 파일의 데이터베이스 색인이 주어진 경우 오디오 샘플과 같은 미디어 샘플을 인식하는 방법에 의해 달성된다. 데이터베이스 색인은 상기 색인이 붙여진 미디어 파일의 특정 위치의 특징을 나타내는 핑거프린트(fingerprint)를 포함한다. 알려지지 않은 미디어 샘플이 데이터베이스내의 미디어 파일(위닝(winning) 미디어 파일)과 동일한 것으로 간주되는데, 이것의 핑거프린트의 상대적인 위치는 샘플의 핑거프린트의 상대적인 위치에 가장 가깝게 매칭된다. 오디오 파일의 경우, 상기 위닝 파일의 핑거프린트의 시간 변화가 샘플의 핑거프린트의 시간 변화와 매칭된다.

상기 방법은 분산 컴퓨터 시스템에서 바람직하게 실행되고, 상기 방법은 샘플의 특정 위치에서의 한 세트의 핑거프린트를 결정하는 단계, 데이터베이스 색인 내에 매칭 핑거프린트를 발견하는 단계, 샘플의 위치와 동등한 핑거프린트를 갖는 파일의 위치 사이의 대응관계를 생성하는 단계, 및 상당히 많은 수의 대응 관계가 실질적으로 선형적으로 관련되는 미디어 파일을 식별하는 단계를 포함한다. 상기 많은 수의 선형적으로 관련되는 대응관계를 갖는 파일이 위닝 미디어 파일로 간주된다. 많은 수의 대응 관계를 갖는 파일을 식별하는 하나의 방법은 대응 관계의 쌍으로부터 생성된 산점도의 대각선에 대한 스케닝에 상당하는 것을 실행하는 것이다. 한 실시예에서, 많은 수의 선형 대응 관계를 갖는 미디어 파일을 식별하는 것은 상기 미디어 파일의 제 1 서브세트만을 검색하는 것을 수반한다. 제 1 서브세트 내의 파일은 상기 제 1 서브세트에 있지 않은 파일보다 식별될 확률이 높다. 식별 확률은 식별 주파수의 선형적 투영과 함께 이전 식별의 새로운 측정 또는 실험에 의한 주파수에 기초하는 것이 바람직하다. 상기 제 1 서브세트에서 미디어 파일이 식별되지 않는 경우, 나머지 파일들을 담고 있는 제 2 서브세트가 검색된다. 대안으로, 상기 파일들은 확률에 의해 분류되고, 상기 분류에 따라 검색된다. 검색은 파일이 발견될 때 종료된다.

샘플 내의 특정한 위치가 상기 샘플에 따라서 재생적으로 계산되는 것이 바람직하다. 이와 같은 재생 가능하게 계산될 수 있는 위치가 "랜드마크(landmark)"로 불린다. 핑거프린트는 숫자로 표시되는 값인 것이 바람직하다. 한 실시예에서, 각각의 핑거프린트는 각각의 위치에 있거나 상기 위치로부터 약간 오프셋되는 다수의 미디어 샘플의 특징을 나타낸다.

오디오 샘플을 인식하는 방법이 특히 유용하며, 이 경우, 특정 위치가 오디오 샘플 내의 타임포인트이다. 예를 들어, 오디오 샘플의 스펙트럼  $L_p$  놈(norm)의 국부 최대값에서 3개의 타임포인트가 발생한다. 핑거프린트는 오디오 샘플의 임의의 분석에 의해 계산될 수 있으며, 상기 샘플의 시간 연장에 불변인 것이 바람직하다. 핑거프린트의 실례로는 스펙트럼 슬라이스 핑거프린트, 멀티 슬라이스 핑거프린트, LPC 계수, 쉘스트럴 계수 및 스펙트로그램 피크의 주파수 성분이 있다.

본 발명은 또한 상기 방법을 실현하기 위한 시스템을 제공하는데, 상기 시스템은 상기 특정 위치를 계산하기 위한 랜드마크 객체(object), 핑거프린트를 계산하기 위한 핑거프린팅 객체, 상기 미디어 파일에 대한 파일 위치와 핑거프린트를 포함하는 데이터베이스 색인 및 분석 객체를 포함한다. 상기 분석 객체는 데이터베이스 내의 매칭 핑거프린트를 발견하고, 대응 관계를 생성하며, 상기 대응 관계를 분석하여 위닝 미디어 파일을 선택함으로써 상기 방법을 실행한다.

또한, 본 발명은 컴퓨터에 의한 액세스가 가능한 프로그램 기억 장치를 제공하며, 상기 프로그램 기억 장치는 상기 컴퓨터에 의해 실행될 수 있는 명령 프로그램을 명확히 구체화하여 상기 방법에 대한 방법의 실행 단계들을 수행한다.

또한, 본 발명은 다수의 오디오 파일 색인을 데이터베이스에 생성하는 방법을 제공하며, 상기 방법은 각각의 파일의 특정 위치에서의 한

세트의 핑거프린트를 계산하는 단계, 파일의 핑거프린트, 위치, 식별자를 메모리 내에 저장하는 단계를 포함한다. 대응하는 핑거프린트, 위치 및 식별자는 트리플렛을 형성하도록 상기 메모리에서 관련된다. 오디오 파일 내의 타임포인트일 수 있는 상기 위치는 파일에 따라서 계산되어 재생될 수 있는 것이 바람직하다. 예를 들어, 타임포인트는 오디오 파일의 스펙트럼 Lp 높의 극대에서 발생할 수 있다. 일부의 경우, 숫자의 값인 것이 바람직한 각각의 핑거프린트는 특정 위치 근방의 파일의 다수의 특징을 나타낸다. 핑거프린트는 오디오 파일의 디지털 신호 처리 또는 임의의 분석으로부터 계산될 수 있다. 핑거프린트의 실례로는 스펙트럼 슬라이스 핑거프린트, 멀티 슬라이스 핑거프린트, LPC 계수, 셉스트럴 계수, 스펙트로그램 피크의 주파수 성분 및 연결된 스펙트로그램 피크가 있다.

결국, 본 발명은 시간 연장 불변의 핑거프린트와 다양한 계층적 검색을 일체화한 오디오 샘플 식별 방법을 제공한다.

#### 도면의 간단한 설명

- 도 1은 사운드 샘플을 인식하는 본 발명의 방법에 대한 흐름도,
- 도 2는 도 1의 방법을 실행하기 위한 예시적인 분산 컴퓨터 시스템의 블록도,
- 도 3은 도 1의 방법에 사용되는 사운드 파일의 데이터베이스 색인을 구성하는 방법의 흐름도,
- 도 4는 사운드 샘플에 대해 계산된 랜드마크와 핑거프린트를 개략적으로 나타내는 도면,
- 도 5는 사운드 샘플에 대한 L4 높의 그래프로써, 랜드마크의 선택을 나타내는 도면,
- 도 6은 도 1의 방법에 사용된 사운드 파일의 데이터베이스 색인을 구성하기 위한 다른 실시예의 흐름도,
- 도 7A-도 7C는 돌출점 및 연결된 돌출점이 표시된 스펙트로그램,
- 도 8A-도 8C는 도 3의 방법의 마스터 색인 리스트, 색인 세트, 색인 리스트를 나타내는 도면,
- 도 9A-도 9C는 도 1의 방법의 색인 리스트, 후보 리스트, 스케터 리스트를 나타내는 도면, 및
- 도 10A-도 10C는 알려지지 않은 사운드 샘플의 정확한 식별 및 식별의 결핍을 각각 나타내는 산점도이다.

#### 실시예

본 발명은 다수의 알려진 미디어 파일을 포함하는 데이터베이스에 제공된 외인성의 미디어 샘플을 인식하는 방법이 제공된다. 본 발명은 또한 본 발명의 인식 방법을 사용하여 효과적인 검색을 허용하는 데이터베이스 색인을 생성하기 위한 방법을 제공한다. 하기의 논의는 주로 오디오 데이터에 대해 언급하지만, 본 발명의 방법은 텍스트, 오디오, 비디오, 이미지 및 개별적인 미디어 형태의 어떠한 멀티미디어의 결합도 포함하지만 이것에 한정되는 것은 모든 종류의 미디어 파일 및 미디어 샘플에 적용될 수 있다는 것을 이해할 수 있을 것이다. 오디오의 경우, 본 발명은 예를 들어 배경 잡음, 전송 에러와 드롭아웃, 대역이 제한된 필터링, 양자화, 시간 왜곡, 및 음질의 디지털 압축에 의해 발생하는 높은 레벨의 선형 및 비선형 왜곡을 포함하는 샘플들을 인식하는데 특히 유용하다. 하기의 상세한 설명으로부터 명백해지는 바와 같이, 본 발명은 이와 같은 조건 하에서 효과가 있는데, 그 이유는 계산된 특징의 일부분만이 왜곡을 견디는 경우라도 왜곡된 신호를 정확히 인식할 수 있기 때문이다. 소리, 음성, 음악 또는 이들의 결합을 포함하는 모든 종류의 오디오가 본 발명에 의해 인식될 수 있다. 오디오 샘플의 실례로는 레코딩된 음악, 라디오 방송 프로그램 및 광고 등이 있다.

본 명세서에서 사용되는 바와 같이, 외인성의 미디어 샘플은 하기에 설명되는 바와 같은 다양한 공급원으로부터 얻어진 임의의 크기의 미디어 데이터의 세그먼트이다. 인식을 실행하기 위해, 상기 샘플은 본 발명에 의해 인덱싱된 미디어 파일의 일부분의 연주(rendering)이어야 한다. 인덱싱된 미디어 파일은 원래의 레코딩으로서 간주될 수 있고, 샘플은 원래의 레코딩의 왜곡 및/또는 요약된 형태 또는 연주로서 간주될 수 있다. 통상적으로, 샘플은 인덱싱된 파일의 일부분에만 대응한다. 예를 들어, 데이터베이스 내에 인덱싱된 5분짜리 노래 중 10초 부분에 대해 인식이 실행될 수 있다. 인덱싱된 엔티티를 설명하기 위해 "파일"이라는 용어가 사용되었지만, 상기 엔티티는 필요한 값(나중에 설명함)이 얻어질 수 있는 임의의 포맷일 수 있다. 또한, 상기 값이 얻어진 후에 파일을 저장하거나 파일에 액세스할 필요가 없다.

도 1에는 본 발명의 방법(10)의 전체 단계를 개략적으로 설명하는 블록도가 도시되어 있다. 개별적인 단계는 하기에 상세히 설명된다. 상기 방법은 워닝 미디어 파일을 식별하고, 미디어 파일의 특징적인 핑거프린트의 상대적인 위치는 외인성의 샘플의 동일한 핑거프린트의 상대적인 위치와 아주 밀접하게 매칭된다. 외인성 샘플이 단계(12)에서 포획된 이후, 랜드마크와 핑거프린트가 단계(14)에서 계산된다. 랜드마크는 샘플 내의 특정 위치, 예를 들어 타임포인트에서 발생된다. 랜드마크의 샘플 내의 위치는 샘플 자체에 의해 결정되는 것이 바람직하다, 즉, 샘플의 특성에 의존하고, 재생 가능한 것이 바람직하다. 즉, 동일한 랜드마크는 프로세스가 반복될 때마다 동일한 신호에 대해 계산된다. 각각의 랜드마크에 대하여, 랜드마크의 근방 또는 랜드마크에서 샘플의 하나 이상의 특징을 특성화하는 핑거프린트가 얻어진다. 랜드마크에 대한 특징의 근접성은 사용된 핑거프린팅 방법에 의해 정의된다. 일부의 경우, 특징이 랜드마크에 명확히 대응하지만, 이전의 또는 이후의 랜드마크에 대응하지 않는 경우, 특징은 랜드마크 근방에 있는 것으로 간주된다. 다른 경우, 특징은 다수의 인접한 랜드마크에 대응한다. 예를 들어, 텍스트 핑거프린트는 문자열일 수 있고, 오디오 핑거프린트는 스펙트럼 성분일 수 있으며, 이미지 핑거프린트는 화소 RGB 값일 수 있다. 단계(14)의 2개의 일반적인 실시예가 하기에 설명되는데, 하나는 랜드마크와 핑거프린트가 순차적으로 계산되는 것이고, 다른 하나는 이들이 동시에 계산되는 것이다.

단계(16)에서, 데이터베이스 색인(18)에 저장된 매칭 핑거프린트의 세트들을 검색하기 위하여 샘플 핑거프린트가 사용되는데, 데이터베이스 색인 내에서 매칭 핑거프린트는 한 세트의 미디어 파일의 식별자 및 랜드마크와 관련된다. 그 다음, 샘플 랜드마크(단계(14)에서 계산됨) 및 동일한 핑거프린트가 계산되어 있는 파일 랜드마크를 포함하는 대응 쌍(단계(20))을 생성하기 위하여, 검색된 파일 식별자와 랜드마크 값의 세트가 사용된다. 그 다음, 그 결과에 따른 대응 쌍이 노래 식별자에 의해 정렬되어, 각각의 적용 가능한 파일에 대한 파일 랜드마크와 샘플 랜드마크 사이의 대응 관계의 세트가 생성된다. 파일 랜드마크와 샘플 랜드마크 사이의 정렬을 위해 각각의 세트가 스캐닝된다. 즉, 랜드마크 쌍의 선형 대응 관계가 식별되고, 선형적으로 관련되는 쌍의 수에 따라서 상기 세트가 기록된다. 허용 한계 내에서, 많은 수의 대응하는 샘플 위치와 파일 위치가 실질적으로 동일한 선형 방정식으로 설명될 수 있을 때, 선형 대응이 발생한다. 예를 들어, 한

세트의 대응 쌍을 설명하는 다수의 방정식의 기울기가  $\pm 5\%$  정도 변하는 경우, 대응 관계의 전체 세트는 선형적으로 관련되는 것으로 간주된다. 물론, 임의의 적절한 허용 한계가 선택될 수 있다. 가장 높은 스코어, 즉 매우 많은 수의 선형으로 관련되는 대응 관계를 갖는 세트의 식별자는 단계(22)에서 발견되고 반환되는 워닝 파일 식별자이다.

아래에 추가로 설명되는 바와 같이, 데이터베이스 내의 엔트리의 수의 로그값에 비례하는 시간 성분에 의해 인식이 실행될 수 있다. 또한 매우 큰 데이터베이스에 대해서도 본질적으로 실시간으로 인식이 실행될 수 있다. 즉, 약간의 시간 지체로 샘플이 얻어질 때 샘플이 인식될 수 있다. 상기 방법은 5-10초 정도 또는 1-3초의 짧은 시간에 기초하여 사운드를 식별할 수 있다. 바람직한 실시예에서, 샘플이 단계(12)에서 포획될 때 랜드마킹과 핑거프린팅 분석(단계(14))이 실시간으로 실행된다. 데이터베이스 조회(단계(16))는 샘플 핑거프린트가 이용 가능하게 될 때 실행되고, 그 대응 결과가 축적되고 선형 대응 관계에 대해 주기적으로 스캐닝된다. 따라서 상기 모든 방법의 단계가 도 1에 제안된 순차적인 선형 방법으로는 아니라 동시에 일어난다. 상기 방법은 텍스트 검색 엔진과 부분적으로 동일하는데, 사용자가 조회 샘플을 제공하고, 사운드 데이터베이스에 인덱싱된 매칭 파일이 반환되는 점에서 동일하다.

상기 방법은 통상적으로 컴퓨터 시스템에서의 소프트웨어 실행으로 수행되며, 상기 방법의 개별 단계들은 독립된 소프트웨어 모듈로서 가장 효율적으로 실행된다. 따라서 본 발명을 실행하는 시스템은 랜드마킹과 핑거프린팅 객체, 인덱싱된 데이터베이스 및 데이터베이스 색인의 검색, 대응 관계의 계산, 워닝 파일의 식별을 위한 분석 객체로 구성되는 것으로 생각될 수 있다. 순차적인 랜드마킹과 핑거프린팅의 경우, 상기 랜드마킹과 핑거프린팅 객체는 다른 랜드마킹과 핑거프린팅 객체인 것으로 간주될 수 있다. 다른 객체에 대한 컴퓨터 명령 코드가 하나 이상의 컴퓨터의 메모리에 저장되고, 하나 이상의 컴퓨터 프로세서에 의해 실행된다. 한 실시예에서, 코드 객체는 인텔 기반 개인 컴퓨터 또는 다른 워크스테이션과 같은 단일 컴퓨터 시스템에서 함께 군집된다. 바람직한 실시예에서, 상기 방법은 중앙 처리 장치(CPU)의 네트워킹된 클러스터에 의해 실행되는데, CPU에서는 계산 가능한 로드를 분산하기 위하여 다른 소프트웨어 객체가 다른 프로세스에 의해 실행된다. 대안으로, 동일하게 구성된 소자의 동종의 네트워킹을 고려하면, 각각의 CPU는 모든 소프트웨어 객체의 복제본을 가질 수 있다. 이러한 후자의 구성에서, 각각의 CPU는 데이터베이스 색인의 서브세트를 갖고, 그 자신의 미디어 파일의 서브세트를 검색할 책임을 갖는다.

본 발명은 임의의 특정한 하드웨어 시스템에 제한되지 않지만, 분산 컴퓨터 시스템(30)의 바람직한 실시예의 실례가 도 2에 개략적으로 설명된다. 시스템(30)은 다중 처리 버스 구조(34)에 의해 연결된 리눅스 기반 프로세서(32a-32f)의 클러스터 또는 Beowulf 클러스터 컴퓨팅 프로토콜과 같은 네트워크 프로토콜, 또는 이들의 혼합체를 포함한다. 이와 같은 구조에서, 데이터베이스 색인은 클러스터의 적어도 하나의 노드(32a)의 RAM(random access memory)에 저장되어, 핑거프린트 검색이 매우 빠르게 이루어지도록 하는 것이 바람직하다. 랜드마킹 노드(32c, 32f), 핑거프린팅 노드(32b, 32e) 및 정렬 스캐닝 노드(32d)와 같은 다른 객체에 대응하는 계산 가능한 노드들은 데이터베이스 색인을 지원하는 노드 또는 노드들(32a)만큼 많은 벌크 RAM을 필요로 하지 않는다. 따라서 각각의 대상에 할당된 계산 가능한 노드의 수는 필요에 따라서 증감되어, 하나의 객체에 병목 현상이 생기지 않도록 할 수 있다. 그러므로 계산 가능한 네트워크는 상당히 병렬화될 수 있고, 이용 가능한 계산 가능 공급원 사이에 분산된 다수의 동시 신호 인식 조회를 추가로 처리할 수 있다. 이것은 많은 수의 사용자가 인식을 요청하고 거의 실시간으로 그 결과를 수신하는 응용을 가능하게 한다는 것에 주목하자.

대안적인 실시예에서, 임의의 기능적인 객체가 매우 엄격하게 서로 결합되는 반면, 나머지는 다른 객체에 엄격하게 결합되지 않는다. 예를 들어, 랜드마킹과 핑거프린팅 객체는 나머지 계산 가능한 객체로부터 물리적으로 분리된 위치에 상주할 수 있다. 이것의 하나의 실례가 랜드마킹과 핑거프린팅 객체의 신호 포획 프로세스와의 밀접한 관련이다. 이 구조에서, 랜드마킹과 핑거프린팅 객체는 예를 들어 이동 전화, WAP(Wireless Application Protocol) 브라우저, PDA(personal digital assistant), 또는 오디오 검색 엔진의 클라이언트 단말기와 같은 다른 원격 단말기에 내장된 추가적인 하드웨어 또는 소프트웨어로서 일체화될 수 있다. 콘텐츠 식별 서비스와 같은 인터넷 기반 오디오 검색 서비스에서, 랜드마킹과 핑거프린팅 객체는 마이크로소프트 DLL(dynamic link library)와 같은 소프트웨어 플러그-인 모듈 또는 링크된 소프트웨어 명령 세트로서 클라이언트 브라우저 애플리케이션에 일체화될 수 있다. 이들 실시예에서, 결합된 신호 포획, 랜드마킹 및 핑거프린팅 객체는 서비스의 클라이언트 단말을 구성한다. 클라이언트 단말은 랜드마크와 핑거프린트 쌍을 포함하는 포획된 신호 샘플의 특성 추출 요약물 서버 말단으로 송신하고, 서버가 인식을 실행한다. 원시 포획 신호(raw captured signal) 대신에 이 특성 추출 요약물 서버로 송신하는 것은 데이터의 양이 500 이상의 인자 정도로 상당히 감소되어 이롭다. 이와 같은 정보는, 예를 들어 서버로 전송되는 오디오 스트림과 함께 또는 이것 대신에 저대역폭 측 채널을 통해서 실시간으로 송신될 수 있다. 이것은 비교적 작은 크기의 대역폭을 각각의 사용자에게 공급하는 공중 통신 네트워크를 통해서 본 발명을 실행하는 것을 가능하게 한다.

사운드 데이터베이스 내에 인덱싱된 오디오 파일과 오디오 샘플을 참조하여 상기 방법이 상세히 설명될 것이다. 상기 방법은 2개의 광범위한 성분, 즉 사운드 데이터베이스 색인 구성과 샘플 인식으로 구성된다.

#### 데이터베이스 색인 구성

사운드 인식이 실행되기 전에, 검색 가능한 사운드 데이터베이스 색인이 구성되어야 한다. 본 명세서에서 사용되는 바와 같이, 데이터베이스는 임의의 데이터의 인덱싱된 집합이고, 상업적으로 이용 가능한 데이터베이스에 제한되지는 않는다. 데이터베이스 색인에서, 관련된 데이터의 요소가 서로 관련되고, 개별적인 요소들이 관련 데이터를 검색하기 위해 사용될 수 있다. 사운드 데이터베이스 색인은 레코딩의 라이브러리 또는 선택된 집합의 각각의 파일 또는 레코딩에 대한 색인 세트를 포함하고, 이것은 음성, 음악, 광고, 소나 기호(sonar signature) 또는 다른 소리를 포함할 수도 있다. 또한, 각각의 레코딩은 고유한 식별자, 즉 sound\_ID를 갖는다. 사운드 데이터베이스 자체는 각각의 레코딩에 대한 오디오 파일을 반드시 저장할 필요는 없지만, 다른 곳으로부터 오디오 파일을 검색하기 위해 sound\_ID가 사용될 수 있다. 사운드 데이터베이스 색인은 매우 커서, 수백만 또는 수십억 개의 파일에 대한 색인을 포함할 것으로 생각된다. 새로운 레코딩이 데이터베이스 색인에 증분으로서 추가되는 것이 바람직하다.

제 1 실시예에 따라서 검색 가능한 사운드 데이터베이스 색인을 구성하기 위한 바람직한 방법(40)의 블록도가 도 3에 도시되어 있다. 이 실시예에서, 랜드마크가 먼저 계산되고, 그 다음 핑거프린트가 랜드마크의 근방 또는 그곳에서 계산된다. 당 분야의 통상의 지식을 가진

자에게 명백한 바와 같이, 데이터베이스 색인을 구성하기 위한 대안적인 방법이 고안될 수 있다. 특히, 하기의 많은 단계들은 선택 사항이지만, 보다 효율적으로 검색되는 데이터베이스를 생성하는데 도움이 된다. 검색 효율이 대용량의 데이터베이스로부터의 실시간 사운드 인식에 중요한 반면, 소용량의 데이터베이스들은 그들이 최적으로 저장되어 있지 않아도 비교적 빠르게 검색될 수 있다.

사운드 데이터베이스를 인덱싱하기 위해, 수집물 내의 각각의 레코딩은 각각의 오디오 파일에 대한 색인을 생성하는 랜드마크킹과 핑거프린팅 분석을 받게 된다. 도 4는 랜드마크(LM)와 핑거프린트(FP)가 계산된 사운드 레코딩의 일부를 개략적으로 나타낸다. 랜드마크는 사운드의 특정 포인트에서 발생하고, 파일의 시작으로부터 오프셋된 시간 단위의 값을 갖고, 핑거프린트는 특정 랜드마크 근방 또는 그곳에서의 사운드의 특징을 나타낸다. 따라서 이 실시예에서, 특정 파일에 대한 각각의 랜드마크는 고유하고, 동일한 핑거프린트가 단일 파일 또는 다수의 파일 내에서 여러 번 발생될 수 있다.

단계(42)에서, 각각의 사운드 레코딩은 상기 사운드 레코딩 내의 독특하고 재생 가능한 위치를 찾기 위한 방법을 사용하여 랜드마크킹된다. 바람직한 랜드마크킹 알고리즘은 잡음 및 다른 선형과 비선형 왜곡의 존재에도 불구하고 사운드 레코딩 내에 동일한 타임포인트를 마킹할 수 있어야 한다. 몇몇 랜드마크킹 방법은 아래에 설명되는 핑거프린트 프로세스와는 개념적으로 무관하지만, 후자의 성능을 최적화하기 위해 선택될 수 있다. 랜드마크킹에 의해 핑거프린트가 차후에 계산되는 사운드 레코딩 내의 타임포인트의 리스트  $\{landmark_k\}$ 가 생긴다. 좋은 랜드마크킹 방법은 사운드 레코딩의 초당 약 5-10개의 랜드마크를 마킹한다. 물론, 랜드마크킹 밀도는 사운드 레코딩 내의 활성도의 양에 따라 달라진다.

랜드마크를 계산하기 위한 다양한 방법이 가능한데, 그 모든 방법들이 본 발명의 범위에 포함한다. 본 발명의 랜드마크킹 방법을 실행하는데 사용되는 특정 기술적 프로세스가 당 분야에 공지되어 있으므로, 상세히 설명되지는 않을 것이다. Power Norm으로 알려진 간단한 랜드마크킹 기술은 레코딩의 모든 가능한 타임포인트에서 순시 전력을 계산하고 극대값을 선택하는 것이다. 이것을 실행하기 위한 하나의 방법은 파형을 직접 정류하고 필터링함으로써 엔벨로프를 계산하는 것이다. 다른 방법은 신호의 Hilbert 변환(구적법(quadrature))을 계산하고, Hilbert 변환의 제곱의 크기와 원 신호의 크기의 합을 이용하는 것이다.

랜드마킹의 Power Norm 방법은 사운드 신호의 과도현상을 찾는데 도움이 된다. 상기 Power Norm은 실제로  $p=2$ 인 아주 일반적인 스펙트럼  $L_p$  놈(Spectral  $L_p$  Norm)의 특별한 경우이다. 일반적인 스펙트럼  $L_p$  놈은 사운드 신호를 따른 각각의 시간에서 Hanning Windowed FFT(Fast Fourier Transform)를 통해서 단시간 스펙트럼을 계산함으로써 계산된다. 바람직한 실시예는 8000Hz의 샘플링 비율, 1024개 샘플의 FFT 프레임 사이즈, 각각의 시간 슬라이스에 대한 64개 샘플의 스트라이드를 사용한다. 그 다음, 각각의 시간 슬라이스에 대한  $L_p$  놈이 스펙트럼 성분의 절대값의  $p^{\text{th}}$  제곱의 합으로써 계산되고, 그 다음에 선택적으로  $p^{\text{th}}$  루트가 취해진다. 상기한 바와 같이, 랜드마크가 시간에 대한 그 결과에 따른 값의 극대값으로서 선택된다. 스펙트럼  $L_p$  놈 방법의 한 실례가 도 5에 도시되며, 특정한 사운드 신호에 대한 시간의 함수로서의  $L_4$  놈의 그래프이다. 극대값에 있는 점선은 선택된 랜드마크의 위치를 나타낸다.

$p = \infty$  인 경우,  $L_{\infty}$  놈이 실제로 최대 놈이다. 즉, 놈의 값은 스펙트럼 슬라이스에서의 가장 큰 스펙트럼 성분의 절대값이다. 이 놈에 의해 강건한 랜드마크 및 양호한 전체 인식 성능이 이루어지므로, 음색의 음악에 좋다.

대안으로, "멀티-슬라이스(multi-slice)" 스펙트럼 랜드마크는 서로로부터 고정 또는 가변의 오프셋에서 스펙트럼 성분의 절대값의  $p^{\text{th}}$  거듭제곱을, 단일 슬라이스 대신에 다수의 슬라이스로 나눈 값의 합을 취함으로써 계산될 수 있다. 이 확장된 합의 극대값을 찾는 것으로 아래에 설명되는 바와 같이 멀티-슬라이스 핑거프린트의 배치의 최적화가 가능하게 된다.

랜드마크가 계산되면, 단계(14)에서 레코딩의 각각의 랜드마크 타임포인트에서 핑거프린트가 계산된다. 핑거프린트는 일반적으로 타임포인트 근방 또는 그곳에서 레코딩에 있는 한 세트의 특징을 요약하는 값 또는 값의 세트이다. 현재의 바람직한 실시예에서, 각각의 핑거프린트는 다수의 특징의 해시 함수(hashed function)인 하나의 수의 값이다. 핑거프린트의 가능한 타입으로는 스펙트럼 슬라이스 핑거프린트, 멀티 슬라이스 핑거프린트, LPC 계수 및 셉스트럴 계수 등이 있다. 물론, 랜드마크 근방의 신호나 신호 특성의 특징을 결정하는 임의의 타입의 핑거프린트도 본 발명의 범위에 속한다. 핑거프린트는 임의의 타입의 디지털 신호 처리 또는 신호의 주파수 분석에 의해 계산될 수 있다.

스펙트럼 슬라이스 핑거프린트를 생성하기 위해, 각각의 랜드마크 타임포인트의 근방에서 주파수 분석이 실행되어 최상의 여러 스펙트럼 피크를 추출한다. 간단한 핑거프린트 값은 단지 가장 강한 스펙트럼 피크의 하나의 주파수 값이다. 이와 같은 단순한 피크의 사용으로 노이즈가 있는 경우에도 매우 양호한 인식이 이루어진다. 그러나 단일 주파수의 스펙트럼 슬라이스 핑거프린트들은 이들이 고유하지 않기 때문에 다른 핑거프린팅 방법보다 많은 폴스 포지티브(false positive)를 생성하는 경향이 있다. 폴스 포지티브의 수는 2개 또는 3개의 가장 강한 스펙트럼 피크의 함수로 이루어진 핑거프린트를 사용함으로써 감소될 수 있다. 그러나 두 번째로 강한 스펙트럼 피크가 잡음이 있는 경우 그의 경쟁자들로부터 잡음을 구별할 수 있을 만큼 충분히 강하지 않은 경우, 잡음에 대해 보다 더 민감할 수 있다. 즉, 계산된 핑거프린트 값은 확실히 재생 가능할 만큼 충분히 강건하지 않을 수 있다. 이것에도 불구하고, 이 경우의 성능 또한 양호하다.

많은 사운드의 시간 변화를 이용하기 위하여, 한 세트의 시간 오프셋을 랜드마크 타임포인트에 추가함으로써 한 세트의 타임슬라이스가 결정된다. 그 결과에 따른 각각의 타임슬라이스에서, 스펙트럼 슬라이스 핑거프린트가 계산된다. 그 다음, 그 결과에 따른 핑거프린트 정보의 세트가 하나의 멀티톤(multitone) 또는 멀티 슬라이스 핑거프린트를 형성하기 위해 결합된다. 각각의 멀티 슬라이스 핑거프린트는 일시적인 변화를 추적하기 때문에 단일한 스펙트럼 슬라이스 핑거프린트보다 훨씬 독특하여, 아래에 설명되는 데이터베이스 색인 검색에서 보다 적은 폴스 매치가 이루어진다. 실험은 이들의 증가된 고유성 때문에 2개의 타임슬라이스의 각각에 있는 하나의 가장 강한 스펙트럼 피크로부터 계산된 멀티 슬라이스 핑거프린트로 인해 이후의 데이터베이스 색인 검색에서 보다 빠르게 계산(약 100배 빠름)되지만, 중대한 잡음이 있는 경우 인식률이 다소 약화된다는 것을 나타낸다.

대안으로, 주어진 타임슬라이스로부터 일정한 오프셋 또는 오프셋들을 사용하여 멀티 슬라이스 핑거프린트를 계산하는 대신, 가변의 오프셋들이 사용될 수 있다. 선택된 슬라이스에 대한 가변 오프셋은 핑거프린트에 대한 "앵커(anchor)" 랜드마크로부터의 임의의 오프셋 범위 내의 랜드마크, 또는 그 다음의 랜드마크에 대한 오프셋이다. 이 경우, 랜드마크들 사이의 타임슬라이스가 또한 다중 주파수 정보와 함께

핑거프린트로 인코딩된다. 보다 많은 차원을 핑거프린트에 추가함으로써, 이들이 보다 고유하게 되어 잘못된 매치가 일어날 가능성이 적어진다.

스펙트럼 성분에 추가해서, 다른 스펙트럼 특징이 추출되어 핑거프린트로서 사용될 수 있다. LPC(linear predictive coding) 분석은 스펙트럼 피크와 같은 신호의 선형적으로 예상 가능한 특성뿐만 아니라 스펙트럼 특성을 추출한다. LPC는 디지털 신호 처리 분야에서 잘 알려져 있다. 본 발명에 대해, 양자화된 LPC 계수를 색인 값으로 해상함으로써 랜드마크 위치에 고정된 파형 슬라이스의 LPC 계수가 핑거프린트로서 사용될 수 있다.

셍스트럴 계수는 주기성의 척도로서 유용하고, 목소리 또는 많은 악기와 같은 음악적인 신호의 특징을 나타내기 위해 사용될 수 있다. 셍스트럴 분석은 디지털 신호 처리의 분야에서 잘 알려져 있다. 본 발명에 대하여, 다수의 셍스트럴 계수가 색인에 함께 해상되고, 핑거프린트로서 사용된다.

랜드마크와 핑거프린트가 동시에 계산되는 다른 실시예(50)가 도 6에 도시되어 있다. 도 3의 단계(42, 44)가 단계(52, 54, 56)로 대체된다. 아래에 설명되는 바와 같이, 다중 차원의 함수가 단계(52)에서 사운드 레코딩으로부터 계산되고, 랜드마크(54)와 핑거프린트(56)가 상기 함수로부터 추출된다.

도 6의 실시예의 하나의 실행에 있어서, 랜드마크와 핑거프린트는 사운드 레코딩의 스펙트로그램으로부터 계산된다. 스펙트로그램은 사운드 샘플의 윈도우 및 중첩 프레임이 스펙트럼에 의해 분석되는 사운드 레코딩의 시간-주파수 분석으로, 통상적으로 FFT(Fast Fourier Transform)를 사용한다. 상기한 바와 같이, 바람직한 실시예는 8000Hz의 샘플링 비율, 1024개 샘플의 FFT 프레임 사이즈, 각각의 타임 슬라이스에 대한 64개 샘플의 스트라이드를 사용한다. 스펙트로그램의 실례가 도 7A에 도시되어 있다. 수평축은 시간이고, 수직축은 주파수이다. 각각의 순차적인 FFT 프레임이 시간축을 따라 대응하는 일정하게 공간을 유지하는 간격으로 수직으로 적층된다. 스펙트럼 도면은 각각의 시간-주파수 포인트에서의 에너지 밀도를 나타내는데, 도면의 어두운 영역이 보다 높은 에너지 밀도를 나타낸다. 스펙트로그램은 오디오 신호 처리의 분야에서 잘 알려져 있다. 본 발명에 대해, 랜드마크와 핑거프린트는 도 7B의 스펙트로그램에서 원으로 표시된, 스펙트로그램의 극대값과 같은 돌출점으로부터 얻어질 수 있다. 예를 들어, 각각의 피크의 시간과 주파수 좌표가 얻어지는데, 시간은 랜드마크에 걸린 시간이고, 주파수는 대응하는 핑거프린트를 계산하는데 사용된 주파수이다. 이 스펙트럼 피크 랜드마크는  $L_\infty$ 놈과 동일하며, 상기 놈의 최대 절대값은 랜드마크 위치를 결정한다. 그러나 스펙트로그램에서, 전체 타임슬라이스보다는 시간-주파수 평면의 패치에 대해 극대값 조사가 실행된다.

이 문맥에서, 사운드 레코딩의 포인트 추출로부터 생긴 돌출점의 세트가 별자리로서 언급된다. 극대값을 구성하는 별자리에 대해, 바람직한 분석은 포인트들을 선택하는 것으로, 상기 포인트들은 각각의 선택된 포인트 주위의 지역에 대한 시간-주파수 평면의 에너지의 최대값이다. 예를 들어, 좌표( $t_j, f_0$ )에서의 포인트가 ( $t_c-T, f_0-F$ ), ( $t_0-T, f_0+F$ ), ( $t_0+T, f_0-F$ ), ( $t_0+T, f_0+F$ )의 코너를 갖는 사각형, 즉, 적절한 별자리 점의 수를 제공하도록 선택된 T와 F를 갖는, 길이(2T와 2F)의 측면을 갖는 사각형 내의 최대 에너지 포인트인 경우, 상기 좌표가 선택된다. 상기 사각형의 경계는 또한 주파수 값에 따라서 그 크기가 변경된다. 물론, 임의의 영역의 형상이 사용될 수 있다. 또한, 경쟁하는 시간-주파수 피크가 시간-주파수 평면의 거리 측정 기준에 따라서 역으로 가중되도록, 즉, 보다 먼 거리의 포인트가 낮은 가중치를 갖도록 최대 에너지 기준이 가중될 수 있다. 예를 들어, 에너지는 다음과 같이 가중될 수 있다.

$$\frac{S(t, f)}{1 + C_t(t - t_0)^2 + C_f(f - f_0)^2}$$

여기서,  $S(t, f)$ 는 포인트( $t, f$ )에서의 스펙트로그램의 값을 제공한 크기이고,  $C_t$ 와  $C_f$ 는 양의 값이다(반드시 일정하지는 않음). 다른 거리-가중 함수도 가능하다. 극대의 선택 상수가 다른(최대가 아님) 돌출점 특성 추출 방법에 적용될 수 있는데, 이것도 본 발명의 범위에 속한다.

이 방법으로 상기한 단일한 주파수 스펙트럼의 핑거프린트와 매우 유사하고, 많은 동일한 특성을 갖는 값의 쌍이 생긴다. 스펙트로그램 시간-주파수 방법은 단일 주파수 방법보다 많은 랜드마크/핑거프린트 쌍을 생성하지만, 아래에 설명되는 매칭 단계에서 많은 잘못된 매치를 산출할 수도 있다. 그러나 이것은 단일 주파수 스펙트럼 핑거프린트보다 강력한 랜드마크링과 핑거프린팅을 제공하는데, 그 이유는 사운드 샘플에서 지배적인 잡음이 각각의 슬라이스의 스펙트럼의 모든 부분으로 확장되지 않을 수도 있기 때문이다. 즉, 지배적인 잡음에 의해 영향을 받지 않는 스펙트럼의 쌍에 몇몇의 랜드마크와 핑거프린트 쌍이 있을 수도 있다.

이 스펙트로그램 랜드마크링과 핑거프린팅 방법은 사운드 신호의 다중 차원(차원 중의 하나가 시간임) 함수를 계산하고 함수의 값에서 돌출점을 찾는 특징 분석 방법의 특별한 경우이다. 돌출점은 극대, 극소, 제로 교차, 또는 다른 특성일 수 있다. 랜드마크가 돌출점의 시간 좌표로 취해지고, 대응하는 핑거프린트가 적어도 나머지 좌표 중 하나로부터 계산된다. 예를 들어, 다중 차원 돌출점의 비시간 좌표(들)가 다중 차원 함수의 핑거프린트를 형성하기 위해 함께 해상될 수 있다.

멀티 슬라이스 스펙트럼 핑거프린트에 대한 상기한 가변의 오프셋 방법은 스펙트로그램 또는 다른 다중 차원 함수 핑거프린트에 적용될 수 있다. 이 경우, 별자리의 포인트들이 도 7C에 도시된 스펙트로그램에 도해된 바와 같이 서로 링크되어 링크 포인트를 형성한다. 별자리의 각각의 포인트는 랜드마크 시간을 정의하는 앵커 포인트로서의 기능을 하고, 다른 포인트들의 나머지 좌표 값은 링크 핑거프린트를 형성하도록 결합된다. 예를 들어, 아래에 정의되는 바와 같이 서로에 인접하는 포인트는 서로 링크되어 보다 쉽게 식별되거나 검색될 수 있는 보다 복잡한 집합 특성 핑거프린트를 형성하도록 서로 결합된다. 멀티 슬라이스 스펙트럼 핑거프린트에서와 같이, 다중 링크 돌출점으로부터의 정보를 단일한 핑거프린트로 결합하는 목적은 보다 다양한 가능한 핑거프린트 값을 생성하는 것이고, 이것에 의해 잘못된 매치의 확률이 감소된다, 즉, 동일한 핑거프린트가 2개의 다른 음악 샘플을 설명할 수 있는 확률이 감소된다.

원칙적으로, N개 돌출점의 각각은 2개의 포인트 연결 방법에서 서로 다른 포인트에 링크되어, 약  $N^2/2$ 개의 결합을 생성한다. 마찬가지로, K-포인트 연결에 대해, 별자리로부터 발생하는 가능한 결합의 수는 약  $N^K$ 개이다. 이와 같은 결합적인 급격한 증가를 회피하기 위해서, 서로 연결되는 점들의 접근을 억제하는 것이 바람직하다. 이와 같은 억제를 달성하기 위한 하나의 방법은 각각의 앵커 포인트에 대한 "목표 지역

(target zone)\*을 정의하는 것이다. 그러면 하나의 앵커 포인트가 그의 목표 지역 내의 포인트들에 링크된다. 연결될 상기 목표 지역 내의 포인트들의 서브세트를 선택하는 것이 가능하다 - 모든 포인트가 링크될 필요는 없다. 예를 들어, 목표 지역의 가장 강한 피크와 관련된 포인트만이 링크될 수 있다. 목표 지역은 일정한 형상을 가질 수 있거나, 앵커 포인트의 특징에 따라서 변할 수 있다. 스펙트로그램 피크 별자리에 대한 앵커 포인트( $t_0, f_0$ )의 목표 지역의 간단한 실례는  $t$ 가 간격  $[t_0+L, t_0+L+W]$ 에 있도록 하는 스펙트로그램 스트립 내의 포인트( $t, f$ )의 세트이고, 여기서  $L$ 은 앞으로의 리드이고,  $W$ 는 목표 지역의 폭이다. 이 방법에서, 모든 주파수가 목표 지역에서 허용된다.  $L$  또는  $W$ 는, 예를 들어 생성될 연결 결합의 수를 조절하기 위하여 비율 제어 메커니즘이 사용되는 경우 변할 수 있다. 대안으로, 예를 들어, 주파수( $f$ )가 간격  $[f_0-F, f_0+F]$ (여기서,  $F$ 는 경계 파라미터) 내에 있도록 목표 지역을 강제함으로써 주파수 제한이 실행될 수 있다. 주파수 억제의 이점은 정신 음향학에서 음표의 시퀀스가 서로 인접한 주파수를 갖는 경우 멜로디가 보다 밀착되는 경향이 있는 것으로 알려져 있다는 점이다. 이와 같은 억제는 음향 정신학을 모델링하는 것이 본 발명의 목표에 반드시 필요한 것은 아니지만 "음향 정신학적으로 사실적인" 인식 성능을 보다 가능하게 할 수 있다. 또한,  $f$ 가 영역  $[f_0-F, f_0+F]$  밖에 있는 반대의 방법을 고려할 수 있다. 이것은 서로 주파수가 다른 점들을 강제로 연결시켜서, 별자리 추출 가공물이 시간적으로 가깝고 동일한 주파수를 갖는 시간-주파수 포인트들의 스퍼터링 시퀀스를 생성하는 경우를 회피할 수 있도록 한다. 다른 위치 파라미터와 마찬가지로,  $F$ 는 반드시 일정할 필요가 없으며, 예를 들어  $f_0$ 의 함수일 수 있다.

핑거프린트 값에 비앵커 돌출점의 시간 좌표를 포함하는 경우, 핑거프린트가 시간 불변이 되도록 상대적인 시간 값이 사용되어야 한다. 예를 들어, 핑거프린트는 (i) 비시간 좌표 값 및/또는 (ii) 돌출점의 대응하는 시간 좌표 값의 차의 함수일 수 있다. 예를 들어, 앵커 포인트에 관하여 또는 링크된 세트 내의 순차적인 돌출점 사이의 연속적인 차이로서 시간 차이가 취해질 수 있다. 해싱된 핑거프린트를 형성하기 위하여 좌표와 차이 값이 연결된 비트 필드로 패킹될 수 있다. 당 분야에 전문적인 지식을 가진 자에게 명백한 바와 같이, 좌표 값의 세트를 핑거프린트 값으로 맵핑하는 많은 다른 방법이 존재하고, 이것도 본 발명의 범위 내에 속한다.

이 방법의 구체적인 실례는 좌표 ( $t_k, f_k$ )(여기서,  $k=1, \dots, N$ )를 갖는  $N > 1$ 개의 링크된 스펙트로그램 피크를 사용한다. 그 다음, (i) 제 1 피크의 시간( $t_1$ )이 랜드마크 시간으로서 취해지고, (ii) 시간 차이( $\Delta t_k = t_k - t_1$ )(여기서,  $k=2, \dots, N$ )에 링크된 피크의 주파수( $f_k$ )(여기서,  $k=1, \dots, N$ )를 더한 값이 서로 해싱되어 핑거프린트 값을 형성한다. 핑거프린트가 모든 이용 가능한  $\Delta t_k$ 와  $f_k$  좌표의 서브세트로부터 또는 모든 좌표로부터 계산될 수 있다. 예를 들어, 시간 차이 좌표의 일부 또는 모두는 원하는 경우 생략될 수 있다.

다수의 포인트를 이용하여 핑거프린트를 형성하는 또 다른 이점은 예를 들어 사운드 레코딩이 원래의 레코딩 속도와는 다른 속도로 다시 실행되는 경우 핑거프린트 인코딩이 시간 연장에 관하여 불변일 수 있다는 것이다. 이러한 이점은 스펙트로그램과 타임슬라이스 방법 모두에 적용된다. 연장된 시간 신호에서, 시간 차이와 주파수는 상호 관계(예를 들어, 두 개의 포인트 사이의 시간 차이를 주파수의 2배만큼 감소시키는 것)를 갖는다는 것에 주목하자. 이 방법은 핑거프린트로부터 시간 연장을 제거하는 방법으로 시간 차이와 주파수를 결합함으로써 상기 사실을 이용한다.

예를 들어, 좌표 값 ( $t_k, f_k$ )(여기서,  $k=1, \dots, N$ )를 갖는  $N$  포인트 스펙트로그램 피크의 경우, 핑거프린트로 해싱하기 위해 이용 가능한 중간 값은  $\Delta t_k = t_k - t_1$ (여기서,  $k=1, \dots, N$ )과  $f_k$ (여기서,  $k=1, \dots, N$ )이다. 그 다음, 상기 중간 값은 상기 주파수들 중의 하나를 기준 주파수, 즉,  $f_1$ 으로서 취함으로써, 그리고 (i) 나머지 주파수와의 몫(quotient)과 (ii) 시간 차이와의 곱을 형성함으로써 시간 연장에 관하여 불변이 될 수 있다.

예를 들어, 중간 값은  $g_k = f_k/f_1$ (여기서,  $k=2, \dots, N$ )과  $s_k = \Delta t_k f_1$ (여기서,  $k=2, \dots, N$ )일 수 있다. 샘플이  $\alpha$ 인자만큼 가속되는 경우, 주파수( $f_k$ )는  $\alpha f_k$ 이 되고, 시간 차이( $\Delta t_k$ )는  $\Delta t_k/\alpha$ 이 되어,  $g_k = \alpha f_k/\alpha f_1 = f_k/f_1$ 과  $s_k = (\Delta t_k/\alpha)(\alpha f_1) = \Delta t_k f_1$ 이 된다. 그 다음, 이들 새로운 중간 값은 함수를 사용하여 결합되어 시간 연장에 독립된 해싱된 핑거프린트를 형성한다. 예를 들어,  $g_k$ 와  $s_k$  값은 이들을 연결된 비트필드로 패킹함으로써 해싱될 수 있다.

대안으로, 기준 주파수 대신, 예를 들어  $\Delta t_2$ 와 같은 기준 시간 차이가 사용될 수 있다. 이 경우, 새로운 중간 값들이 (i) 나머지 시간 차이와의 몫  $\Delta t_k/\Delta t_2$ 와, (ii) 기준 주파수와의 곱  $\Delta t_2 f_k$ 로서 계산된다. 이 경우는 기준 주파수를 사용하는 것과 동등한데, 그 이유는 그 결과의 값이 상기의  $g_k$ 와  $s_k$  값의 곱과 몫으로부터 형성될 수 있기 때문이다. 주파수 비율의 역수가 동일하게 유효하게 사용될 수 있다. 원래 중간 값의 로그값의 합과 차가 또한 곱과 차 대신에 각각 사용될 수 있다. 이와 같은 수학적 연산의 교환, 대입, 치환에 의해 얻어진 시간 연장 독립의 핑거프린트가 본 발명의 범위 내에 포함된다. 또한, 시간 차이를 상대적으로 다루는 기준 시간 차이나 다수의 기준 주파수가 사용될 수 있다. 다수의 기준 주파수나 기준 시간 차이의 사용은 단일 기준의 사용과 동등한데, 그 이유는  $g_k$ 와  $s_k$  값의 계산상의 조작에 의해 동일한 결과가 달성될 수 있기 때문이다.

도 3과 도 6을 참조하면, 상기한 방법 중의 어느 하나에 의한 랜드마크링과 핑거프린팅 분석으로 도 8A에 도시된 바와 같이 각각의 Sound\_ID에 대한 색인 세트가 생긴다. 소정의 사운드 레코딩에 대한 색인 세트는 값의 쌍(fingerprint, landmark)의 리스트이다. 각각의 인덱싱된 레코딩은 통상적으로 그의 색인 세트 내에 약 1000개의 (fingerprint, landmark)의 쌍을 갖는다. 랜드마크링과 핑거프린팅 기술이 본질적으로 독립인 상기한 제 1 실시예에서, 이들은 독립적이고 상호 교환 가능한 모듈로서 취급된다. 상기 시스템, 신호 품질, 또는 인식될 사운드의 형태에 따라서, 다수의 다른 랜드마크링 또는 핑거프린팅 모듈 중의 하나가 사용될 수 있다. 실제로, 색인 세트가 단순히 값의 쌍으로 구성되기 때문에, 다수의 랜드마크링 및 핑거프린팅 방법이 동시에 사용되는 것이 가능하고, 때로는 이것이 바람직하다. 예를 들어, 고유한 음색의 패턴을 검출하는데 하나의 랜드마크링 및 핑거프린팅 방법이 좋을 수 있지만, 진동을 식별하는데는 불충분하기 때문에, 다른 알고리즘이 반대의 속성을 가질 수 있다. 다수의 랜드마크링/핑거프린팅 전략의 사용으로 보다 강력하고 넓은 범위의 인식 성능이 달성된다. 임의의 종류의 핑거프린트에 대해 임의의 범위의 핑거프린트 값을 유지함으로써 다른 핑거프린팅 기술이 함께 사용될 수 있다. 예를 들어, 32 비트 핑거프린트 값에서, 앞의 3비트는 8개 핑거프린팅 계획 중의 어느 것을 특정하기 위해 사용될 수 있고, 나머지 29 비트는 인코딩된다.

사운드 데이터베이스 내에 인덱싱될 각각의 사운드 레코딩에 대해 색인 세트가 생성된 이후, 빠른(즉, 로그 시간) 검색을 허용하기 위한 방법으로 검색 가능한 데이터베이스 색인이 구성된다. 이것은, 대응하는 Sound\_ID를 각각의 색인 세트 내의 각각의 더블릿(doublet)에 첨부함으로써 얻어지는 트리플릿(fingerprint, landmark, sound\_ID)의 리스트를 구성함으로써 단계(46)에서 달성된다. 모든 사운드 레코딩에 대한 이와 같은 모든 트리플릿이 커다란 색인 리스트 내에 모이는데, 이것의 실례가 도 8B에 도시되어 있다. 이후의 검색 프로세스를



최적화하기 위해, 트리플렛의 리스트가 핑거프린트에 따라서 저장된다. 고속 정렬 알고리즘(fast sorting algorithm)은 당 분야에 잘 알려져 있고, D.E. Knuth, The Art of Computer Programming, Volume 3: Sorting and Searching, Reading, Massachusetts: Addison-Wesley, 1998에 광범위하게 개시되어 있고, 이것은 참고로 본 명세서에 일체화되어 있다.  $N \log N$  시간의 리스트를 정렬하기 위하여 고성능 정렬 알고리즘이 사용될 수 있으며, 여기서  $N$ 은 리스트로의 엔트리의 수이다.

색인 리스트가 정렬되고 나면, 상기 리스트 내의 각각의 고유한 핑거프린트가 새로운 마스터 색인 리스트에 모이도록 분할됨으로써 단계 (48)에서 추가로 처리되고, 이것의 실례가 도 8C에 도시되어 있다. 마스터 색인 리스트로의 각각의 엔트리는 핑거프린트 값과 (landmark, sound\_ID) 쌍의 리스트로의 포인터를 포함한다. 인덱싱된 레코딩의 수와 특성에 따라서, 소정의 핑거프린트는 전체 컬렉션 내에서 수백 배 이상인 것으로 보일 수 있다. 색인 리스트를 마스터 색인 리스트 내에 재정렬하는 것은 선택 사항이지만, 비용을 절감할 수 있는데, 그 이유는 각각의 핑거프린트 값이 단 한번만 나타나기 때문이다. 또한, 리스트 내로의 유효한 엔트리의 수가 고유한 값의 리스트로 상당히 감소되기 때문에, 이것은 차후의 데이터베이스 검색을 가속시킨다. 대안으로, 마스터 색인 리스트는 각각의 트리플렛을 B-트리로 삽입함으로써 구성될 수 있다. 당 분야의 전문가들에게 잘 알려진 바와 같이, 마스터 색인 리스트를 구성하는 다른 가능성도 존재한다. 마스터 색인 리스트는 신호 인식 동안 고속 액세스를 위하여 DRAM과 같은 시스템 메모리 내에 유지되는 것이 바람직하다. 마스터 색인 리스트는 도 2에 도시된 바와 같이 시스템 내의 단일 노드의 메모리에 유지될 수 있다. 대안으로, 마스터 색인 리스트는 다수의 계산 가능한 노드 사이에서 분산되는 조각들로 갈라질 수 있다. 상기한 사운드 데이터베이스 색인은 도 8C에 도해진 마스터 색인 리스트인 것이 바람직하다.

사운드 데이터베이스 색인은 오프라인에서 구성되고, 새로운 사운드가 인식 시스템으로 일체화될 때 증분으로써 업데이트되는 것이 바람직하다. 리스트를 업데이트하기 위해, 새로운 핑거프린트가 마스터 리스트의 적절한 위치로 삽입될 수 있다. 새로운 레코딩이 현재의 핑거프린트를 포함하는 경우, 대응하는 (landmark, sound\_ID) 쌍이 이들 핑거프린트들에 대한 현재의 리스트에 추가된다.

### 인식 시스템

상기한 바와 같이 생성된 마스터 색인 리스트를 사용함으로써, 외인성의 사운드 샘플에 대한 사운드 인식이 실행되는데, 외인성의 사운드 샘플은 통상적으로 상기 샘플을 식별하는데 관심이 있는 사용자에게 의해 공급된다. 예를 들어, 사용자는 라디오를 통해서 새로운 노래를 듣고, 그 노래의 곡명과 가수를 알고 싶어 한다. 샘플은 라디오 방송, 디스코 음악, 대중음악, 서브마린, 사운드 파일, 스트리밍 오디오의 세그먼트 또는 스테레오 시스템과 같은 임의의 타입의 환경으로부터 생성될 수 있고, 배경 잡음, 드롭아웃, 또는 말하는 목소리를 포함할 수도 있다. 사용자는 오디오 샘플을 인식 시스템에 제공하기 전에 이것을 자동 응답 장치, 컴퓨터 파일, 테이프 레코더, 또는 전화나 이동 전화 음성 메일 시스템과 같은 기억 장치 내에 저장할 수 있다. 시스템 세트업과 사용자 억제 조건에 기초하여, 오디오 샘플은 스테레오 시스템, 텔레비전, 컴팩트 디스크 플레이어, 라디오 방송, 자동 응답 장치, 전화기, 이동 전화, 인터넷 스트리밍 방송, FTP, 이메일 첨부 자료로서의 컴퓨터 파일, 또는 이와 같은 기록된 자료의 전송을 위한 다른 모든 적절한 수단 등의 임의의 수의 아날로그 또는 디지털 공급원으로부터 본 발명의 인식 시스템에 제공된다. 상기 공급원에 따라서, 샘플은 음파, 전파, 디지털 오디오 PCM 스트림, 압축된 디지털 오디오 스트림(예를 들어, 돌비 디지털(Dolby Digital) 또는 MP3), 또는 인터넷 스트리밍 방송의 형태일 수 있다. 사용자는 전화기, 이동 전화기, 웹 브라우저, 또는 이메일과 같은 표준 인터페이스를 통해 인식 시스템과 상호 작용을 한다. 샘플은 상기 시스템에 의해 캡처되고 실시간으로 처리될 수 있거나, 이전에 캡처된 사운드(예를 들어, 사운드 파일)로부터의 처리를 위해 재생될 수 있다. 캡처되는 동안, 오디오 샘플은 마이크로폰과 같은 샘플링 장치로부터 디지털로 샘플링되어 상기 시스템으로 송신된다. 캡처 방법에 따라서, 샘플은 채널 또는 사운드 캡처 장치의 제한에 기인한 추가적인 저하를 받을 수 있다.

사운드 신호가 디지털 형태로 변환되면, 이것은 인식을 위해 처리된다. 데이터베이스 파일에 대한 색인 리스트의 구성에서와 마찬가지로, 사운드 레코딩 데이터베이스를 처리하기 위하여 사용된 알고리즘과 동일한 알고리즘을 사용하여 샘플에 대한 랜드마크와 핑거프린트가 계산된다. 원래의 사운드 파일의 상당히 왜곡된 연주의 처리가 원래의 레코딩에 대해 얻어진 것과 유사 또는 동일한 랜드마크 및 핑거프린트 쌍을 산출하는 경우, 상기 방법은 최적으로 작용한다. 상기 사운드 샘플에 대한 그 결과의 색인 세트가 도 9A에 도시된 바와 같은 한 세트의 분석된 값의 쌍, 즉, (fingerprint, landmark)이다.

사운드 샘플에 대해 상기 쌍이 주어지는 경우, 매칭 파일을 잠재적으로 찾기 위하여 데이터베이스 색인이 검색된다. 검색은 다음과 같이 실행된다. 알려지지 않은 샘플의 색인 세트 내의 각각의 (fingerprint<sub>k</sub>, landmark<sub>k</sub>) 쌍은 마스터 색인 리스트 내의 fingerprint<sub>k</sub>를 찾음으로써 처리된다. 정렬된 리스트에 대한 고속 검색 알고리즘은 당 분야에 잘 알려져 있고, D.E. Knuth, The Art of Computer Programming, Volume 3: Sorting and Searching, Reading, Massachusetts: Addison-Wesley, 1998에 광범위하게 개시되어 있다. 마스터 색인 리스트에서 fingerprint<sub>k</sub>이 확인되면, 그의 대응하는 매칭 리스트 (landmark<sub>j</sub>, sound\_ID<sub>j</sub>) 쌍이 복제되고, landmark<sub>k</sub>가 보강되어 형태 (landmark<sub>k</sub>, landmark<sub>j</sub>, sound\_ID<sub>j</sub>)의 한 세트의 트리플렛이 형성된다. 이 표기법에서, 별표(\*)는 데이터베이스 내의 인덱싱된 파일들 중의 하나의 랜드마크를 나타내고, 별표가 없는 랜드마크는 샘플로서 언급된다. 몇몇의 경우, 상기 매칭 핑거프린트는 반드시 동일할 필요는 없지만, 동일한 것이 바람직하다. 예를 들어, 이들은 이전에 결정된 임계값의 범위 내에서 다를 수 있다. 유사하거나 동일한 매칭 핑거프린트는 동등한 것으로서 언급된다. 트리플렛의 sound\_ID<sub>j</sub>는 별표의 랜드마크를 갖는 파일에 대응한다. 따라서 각각의 트리플렛은 2개의 다른 랜드마크를 포함하는데, 하나는 데이터베이스 색인 내의 것이고, 다른 하나는 샘플 내의 것인데, 각각의 트리플렛에서 동등한 핑거프린트가 계산된다. 이 프로세스는 입력 샘플 색인 세트의 범위에 걸친 모든  $k$ 에 대해 반복된다. 도 9B에 도시된 바와 같이 그 결과의 모든 트리플렛이 커다란 후보 리스트 내로 모인다. 상기 후보 리스트는 사운드 파일의 sound\_ID를 포함하기 때문에 이들의 매칭 핑거프린트에 의해 외인성의 사운드 샘플과 동일한 후보로 불린다.

상기 후보 리스트가 편집된 이후, 이것은 sound\_ID에 따라서 분할됨으로써 추가로 처리된다. 이것을 실행하는 편리한 방법은 sound\_ID에 의해 상기 후보 리스트를 정렬하거나 이것을 B-트리에 삽입하는 것이다. 상기한 바와 같이, 상당히 많은 수의 정렬 알고리즘이 당 분야에서 이용 가능하다. 이 프로세스의 결과가 후보 sound\_ID들의 리스트이고, 이것의 각각은 도 9C에 도시된 바와 같이 sound\_ID가

선택적으로 빠지는 샘플 및 파일 랜드마크 타임포인트의 쌍의 스케터 리스트, 즉,  $(\text{landmark}_n, \text{landmark}_n')$ 를 갖는다. 따라서 각각의 스케터 리스트는 한 세트의 대응하는 랜드마크를 포함하는데, 이들의 동등한 핑거프린트 값의 특징에 의해 대응된다.

그 다음, 각각의 후보 sound\_ID에 대한 스케터 리스트가 분석되어 상기 sound\_ID가 샘플에 대한 매치인지 여부를 결정된다. 매우 작은 스케터 리스트를 갖는 잠재적으로 많은 수의 후보들을 제거하기 위하여 먼저 선택적인 임계값 결정 단계가 사용된다. 명백하게, 그들의 스케터 리스트로의 단 하나의 엔트리, 즉, 샘플과 같은 단 하나의 핑거프린트를 갖는 후보들은 상기 샘플에 매칭되지 않는다. 1 이상의 적절한 임계값 번호가 사용될 수 있다.

후보들의 최종적인 수가 결정되면, 위닝 후보의 위치가 정해진다. 그 다음의 알고리즘이 위닝 후보의 위치를 발견하지 못하는 경우, 실패 메시지가 반환된다. 매칭 프로세스의 핵심적인 통찰력은 양 쪽의 시간 축이 일정하다고 가정하면 사운드 매칭에서의 시간 변화가 선형의 대응관계를 따라야 한다는 것이다. 사운드 중의 하나가 비선형적으로 계획적으로 왜곡되지 않거나 떨림이 있는 속도 문제를 갖는 테이프 댁과 같은 결함이 있는 플레이백 장치에 영향을 받지 않는 경우에 이것은 거의 항상 틀림없다. 따라서 소정의 sound\_ID의 스케터 리스트 내의 정확한 랜드마크 쌍  $(\text{landmark}_n, \text{landmark}_n')$ 은,

$$\text{landmark}_n' = m \cdot \text{landmark}_n + \text{오프셋}$$

형태의 선형 대응 관계를 가져야 한다. 여기서, m은 기울기로서, 1에 가까워야 하고,  $\text{landmark}_n$ 는 외인성의 샘플 내의 타임포인트이고,  $\text{landmark}_n'$ 는 sound\_ID에 의해 인덱싱된 사운드 레코딩 내의 대응하는 타임포인트이고, 오프셋은 상기 외인성의 사운드 샘플의 시작에 대응하는 사운드 레코딩으로의 시간 오프셋이다. m과 오프셋의 특정값에 대해 상기 식이 적합할 수 있는 랜드마크의 쌍은 선형으로 관련되는 것으로 언급된다. 분명히, 선형으로 관련된다는 개념은 한쌍의 이상의 대응하는 랜드마크에 대해서만 유효하다. 이 선형 관계는 중요성이 없는 멀리 떨어진 랜드마크 쌍을 제외하면서 높은 확률을 가진 정확한 사운드 파일을 식별한다는 것에 주목하자. 두 개의 구별되는 신호가 다수의 동일한 핑거프린트를 포함하는 것이 가능하지만, 이들 핑거프린트들이 동일한 상대 시간 변화를 가질 가능성은 없다. 선형 대응 관계에 대한 요구 조건은 본 발명의 중요한 특징으로, 공통의 특징의 전체 수를 간단히 계수하거나 특성 사이의 유사성을 측정하는 기술보다 상당히 양호한 인식을 제공한다. 실제, 본 발명의 이러한 측면 때문에, 1% 미만의 원래의 레코딩의 핑거프린트가 외인성의 사운드 샘플에 나타내는 경우, 즉, 사운드 샘플이 매우 짧거나 상당히 왜곡되어 있는 경우에도 사운드가 인식될 수 있다.

따라서 외인성 샘플에 대한 매칭이 존재하는지를 결정하는 문제는 소정의 스케터 리스트의 랜드마크 포인트의 산점도 내의 1에 가까운 기울기를 갖는 대각선을 찾는 것에 상당하는 것으로 해결된다. 도 10A와 도 10B에 2개의 샘플 산점도가 도시되어 있는데, 수평축이 사운드 파일 랜드마크이고, 수직축이 외인성의 사운드 샘플 랜드마크이다. 도 10A에서, 거의 기울기가 1에 가까운 대각선이 확인되는데, 이것은 노래가 실제로 샘플에 매칭됨을, 즉, 사운드 파일이 위닝 파일임을 나타낸다. 수평축 절편은 샘플이 시작되는 오디오 파일로의 오프셋을 나타낸다. 도 10B에는 통계적으로 중요한 대각선이 발견되지 않았는데, 이것은 사운드 파일이 외인성의 샘플에 대한 매치가 아님을 나타낸다.

산점도에서 대각선을 찾는 많은 방법이 있는데, 그 모든 방법이 본 발명의 범위 내에 포함된다. "대각선을 찾음"이라는 말은 대각선을 명백히 만들지 않고 대각선을 찾는 것과 동등한 모든 방법을 언급한다는 것을 이해할 수 있을 것이다. 바람직한 방법은 상기 수식식 1의 양변에서  $m \cdot \text{landmark}_n$ 을 감산하여

$$(\text{landmark}_n' - m \cdot \text{landmark}_n) = \text{오프셋}$$

을 산출함으로써 시작된다. m이 1과 거의 같고 시간 연장이 없는 것을 가정하면, 우리는 하기의 식을 얻는다.

$$(\text{landmark}_n' - \text{landmark}_n) = \text{오프셋}$$

그 다음, 대각선 발견 문제는 동일한 오프셋 값 근방에 밀집되어 있는 소정의 sound\_ID에 대한 다수의 랜드마크 쌍을 찾는 것으로 해결된다. 이것은 하나의 랜드마크를 나머지 것에서 감산하고, 그 결과의 오프셋 값의 막대그래프를 수집함으로써 간단히 달성될 수 있다. 고속 정렬 알고리즘을 사용하여 상기 결과의 오프셋 값을 정렬함으로써 또는 카운터로 빈 엔트리(bin entry)를 생성하고 B 트리에 삽입함으로써 막대그래프가 작성될 수 있다. 상기 막대그래프의 위닝 오프셋 빈은 가장 높은 빈도의 포인트를 포함한다. 이 빈은 본 명세서에서 막대그래프의 피크로서 언급된다. 외인성의 사운드 신호가 정확한 라이브러리 사운드 파일 내에 충분히 포함되는 경우 오프셋이 포지티브이어야 하기 때문에, 네가티브 오프셋이 되는 랜드마크 쌍들이 배제될 수 있다. 마찬가지로, 파일의 말단을 넘는 오프셋 또한 배제될 수 있다. 막대그래프의 위닝 오프셋 빈 내의 포인트의 수가 각각의 자격을 갖는 sound\_ID에 대해 기록된다. 이 수는 각각의 사운드 레코딩에 대한 스코어가 된다. 가장 높은 스코어를 갖는 후보 리스트 내의 사운드 레코딩이 위너(winner)로서 선택된다. 위닝 sound\_ID는 아래에 설명되는 바와 같이 식별의 성공을 신호로 전송하기 위해 사용자에게 보고된다. 식별의 실패를 방지하기 위하여, 식별 프로세스의 성공을 게이트로 제어하기 위해 최소 임계 스코어가 사용될 수 있다. 라이브러리 사운드가 상기 임계값을 넘는 스코어를 갖지 못하면, 인식이 되지 않고, 사용자가 통지를 받는다.

외인성의 사운드 신호가 다수의 사운드를 포함하는 경우, 각각의 개별적인 사운드가 인식될 수 있다. 이 경우, 다수의 위너가 정렬 스캔에서 발견된다. 상기 정렬 스캔이 나머지 스코어보다 훨씬 높은 스코어를 갖는 하나 이상의 sound\_ID를 찾을 것이기 때문에, 사운드 신호가 다수의 위너를 포함한다는 것을 알 필요가 없다. 사용된 핑거프린팅 방법은 양호한 선형의 중첩을 나타내므로, 개별적인 핑거프린트가 추출될 수 있다. 예를 들어, 스펙트로그램 핑거프린팅 방법은 선형의 중첩을 나타낸다.

사운드 샘플이 시간 연장을 받은 경우, 기울기는 1과 같지 않다. 시간 연장 샘플에 대해 1이라는 기울기를 가정(핑거프린트가 시간 연장 불변인 것을 가정)한 결과는 계산된 오프셋 값이 동일하지 않다는 것이다. 이것을 설명하고 적절한 시간 연장을 수용하기 위한 하나의 방법은 오프셋 빈의 크기를 증가시키는 것, 즉, 오프셋의 범위가 동일하게 되는 것을 고려하는 것이다. 일반적으로, 상기 포인트들이 직선으로 내려가지 않는다면, 계산된 오프셋 값이 상당히 달라서, 오프셋 빈의 크기의 약간의 증가는 상당한 수의 잘못된 포지티브를 산출하지 않는다.

다른 선 발견 전략도 가능하다. 예를 들어, T. Risse, "Hough Transform for Line Recognition", Computer Vision and Image Processing, 46, 327-345, 1989에 개시되고, 머신 비전과 그래픽 조사의 분야에 잘 알려져 있는 라돈 또는 허프 변환이 사용될 수 있다. 허프 변환에서, 산점도의 각각의 점은 (기울기, 오프셋) 공간의 선에 투영된다. 따라서 산점도 내의 점들의 세트가 허프 변환에서 이중 공간으로

투영된다. 허프 변환의 피크들은 파라미터 라인들의 교차점에 대응한다. 소정의 산점도의 이와 같은 변환의 전역 피크는 허프 변환에서 교차선의 최상의 수, 즉 공직선(co-linear)의 점의 최상의 수를 나타낸다. 5%의 속도 변화를 허용하기 위해, 예를 들어, 허프 변환의 구조는 기울기 파라미터가 0.95와 1.05 사이에서 변화하는 영역에 제한되어 약간의 계산상의 에러를 줄일 수 있다.

#### 계층적 검색

매우 작은 스케터 리스트를 갖는 후보들을 제거하는 역치 단계에 추가해서, 추가로 효율의 개선이 이루어질 수 있다. 이와 같은 하나의 개선안에서, 데이터베이스 색인은 발생 확률에 따라서 적어도 2개의 부분으로 분할되고, 샘플을 매칭할 가장 높은 확률을 갖는 사운드 파일만이 최초로 검색된다. 다수의 프로세스의 단계에서 분할이 일어날 수 있다. 예를 들어, 마스터 색인 리스트(도 8C)가 2개 이상의 부분으로 분할되어 단계(16과 20)가 상기 분할 부분 중의 하나에 대하여 실행된다. 즉, 매칭 핑거프린트에 대응하는 파일들이 데이터베이스 색인의 일부분으로부터 검색되고, 이 부분으로부터 스케터 리스트가 생성된다. 위닝 사운드 파일이 발견되지 않는 경우, 데이터베이스 색인의 나머지에 대해 상기 프로세스가 반복된다. 다른 실시예에서, 모든 파일이 데이터베이스 색인으로부터 검색되지만, 다른 부분에 대해서는 대각선 스캔이 독립적으로 실행된다.

이 기술을 사용하여, 상기 방법의 계산상으로 집중되는 부분인 대각선 스캔이 데이터베이스 색인의 사운드 파일의 작은 서브세트에 대하여 먼저 실행된다. 대각선 스캔은 스캐닝될 사운드의 수에 대해 거의 선형인 시간 성분을 갖기 때문에, 이와 같은 계층적 검색을 실행하는 것이 상당히 이롭다. 예를 들어, 사운드 데이터베이스 색인이 1,000,000개의 사운드 파일을 나타내는 핑거프린트를 포함하지만, 약 1,000개의 파일만이 높은 주파수를 갖는 상기 샘플 조화 및 매칭되는데, 예를 들어, 95%의 조화가 1000개의 파일에 대응하고, 단 5%의 조화만이 나머지 999,000개의 파일에 대응한다고 가정하자. 그러므로 평균 비용은 약 50,900에 비례한다. 따라서 계층적 검색은 계산할 수 있는 부하에서 약 20개의 수익 인자를 산출한다. 물론, 데이터베이스 색인은 2레벨 이상의 계층, 예를 들어 새로운 출시물의 그룹, 최근에 출시된 노래의 그룹, 및 오래되고 대중적이지 못한 노래의 그룹으로 분할될 수 있다.

상기한 바와 같이, 높은 확률 파일인 사운드 파일의 제 1 서브세트에 대해 먼저 검색이 실시되고, 제 1 검색이 실패한 경우에만 나머지 파일을 포함하는 제 2 서브세트에 대해 검색이 실행된다. 각각의 오프셋 빈 내의 포인트의 수가 소정의 임계값에 도달하지 않는 경우 대각선 스캔의 실패가 발생한다. 대안으로, 두 개의 검색이 병렬로(동시에)실행될 수 있다. 제 1 서브세트의 검색에서 정확한 사운드 파일이 발견되는 경우, 제 2 서브세트의 검색을 종료시키는 신호가 송신된다. 상기 제 1 검색에서 정확한 사운드 파일이 발견되지 않는 경우, 위닝 파일이 발견될 때까지 제 2 검색이 계속된다. 이들 2개의 다른 실행은 계산의 노력과 시간의 균형을 수반한다. 첫 번째 실행은 계산상으로 보다 효율적이지만, 제 1 검색이 실패하는 경우 약간의 대기 시간을 유도하고, 두 번째 실행은 위닝 파일이 제 1 서브세트 내에 있는 경우 불필요한 계산상의 노력이 들지만, 위닝 파일일 제 1 서브세트 내에 있지 않은 경우 대기시간을 최소화한다.

리스트를 분할하는 목적은 제 2 파일이 조화의 목표에 있을 확률을 추정하고, 조화 샘플과 가장 매칭될 것 같은 이들 파일에 검색을 제한하는 것이다. 데이터베이스의 사운드를 정렬하고 확률을 할당할 수 있는 많은 가능한 방법이 있는데, 이 모든 것이 본 발명의 범위 내에 포함된다. 확률은 위닝 사운드 파일로서 식별되는 것의 빈도 또는 구매의 최신성(recency)에 기초하여 할당된다. 새로운 노래가 출시됨에 따라 음악적인 관심은 시간이 지날수록 상당히 빠르게 변화하기 때문에 최신성은 특히 대중 가요에 대해 유용한 척도이다. 확률 스코어가 계산된 이후, 순위가 파일에 할당되고, 리스트가 상기 순위에 의해 스스로 정렬된다. 그 다음, 정렬된 리스트는 검색을 위해 2개 이상의 서브세트로 분할된다. 보다 작은 서브세트는 소정 수의 파일을 포함할 수 있다. 예를 들어, 순위가 상위의 파일, 즉, 1000개의 파일 내에 위치하면, 파일은 보다 작고 빠른 검색으로 배열된다. 대안으로, 2개의 서브세트에 대한 차단점은 동적으로 조정될 수 있다. 예를 들어, 특정 임계값을 초과하는 스코어를 갖는 모든 파일이 제 1 서브세트 내에 놓일 수 있어서, 각각의 서브세트 내의 파일의 수가 연속적으로 변화된다.

상기 확률을 계산하는 하나의 구체적인 방법은 사운드 파일의 스코어가 조화 샘플에 대한 매치로서 식별될 때마다 상기 사운드 파일의 스코어를 1씩 증가시키는 것이다. 구매의 최신성을 고려하기 위해, 모든 스코어가 주기적으로 하향으로 감소되어, 새로운 조화가 이전의 조화보다 순위에 대해 강한 효과를 갖는다. 예를 들어, 모든 스코어가 각각의 조화에 대해 일정한 곱셈 인자만큼 서서히 내려갈 수 있게 되어, 스코어가 갱신되지 않는 경우 스코어가 기하급수적으로 감소하게 된다. 쉽게 백만 개가 될 수 있는 데이터베이스 내의 파일의 수에 따라서, 이 방법은 모든 조화에서 많은 수의 스코어의 갱신을 필요로 할 수 있으므로, 바람직하지 않을 수도 있다. 대안으로, 상기 스코어는 비교적 빈번하지 않은 간격, 예를 들어 하루에 한 번의 간격으로 하향으로 조정될 수 있다. 빈번하지 않은 조정으로 생긴 배열은 각각의 조화에 대한 조정으로부터 생긴 배열과 실제로 동일하지만 거의 동일하지는 않다. 그러나 순위를 갱신하기 위한 계산 가능한 부하는 매우 낮다.

최신 스코어를 보다 정확히 보존하는 이 새로운 조정의 약간의 변화는 지수적으로 증가하는 스코어 업데이트( $a^t$ )를 조화의 위닝 사운드 파일에 추가하는 것이며, 여기서  $t$ 는 마지막 전역 업데이트 이후에 경과된 시간의 양이다. 그 다음, 각각의 전역 업데이트에서  $a^T$ 로 나눔으로써 모든 스코어가 조정되며, 여기서  $T$ 는 마지막 전역 업데이트 이후의 최종 경과 시간이다. 이 변수에서,  $a$ 는 1보다 큰 최신성 인자(recency factor)이다.

상기한 순위에 추가해서, 목록 작성의 시드 배정을 돕기 위해서 몇몇 선험적 지식이 도입될 수 있다. 예를 들어, 새로운 출시품이 이전의 노래보다 높은 조화 변동을 가질 수 있다. 따라서 새로운 출시품은 매칭 조화의 높은 확률을 갖는 노래들을 포함하는 제 1 서브세트 내에 자동으로 놓일 수 있다. 이것은 상기한 자체 순위 결정 알고리즘에서 독립적으로 실행될 수 있다. 또한, 자체 순위 결정 특징이 사용되는 경우, 새로운 출시품은 그들을 제 1 서브세트 내의 어딘가에 배치하는 최초 순위에 할당될 수 있다. 새로운 출시품은 높은 확률의 노래의 리스트의 최상위, 상기 리스트의 최하위, 또는 그 사이의 어딘가에 시드를 배정받을 수 있다. 검색을 위해, 관심의 진실 수준을 반영하기 위해 시간에 걸쳐 배열 순위가 집중되기 때문에 최초 위치는 중요하지 않다.

다른 실시예에 있어서, 최신 순위를 위해 검색이 실행되고, sound\_ID 스코어가 소정의 임계값을 초과하는 경우에 검색이 종료된다. 이것은 각각의 세그먼트가 하나의 sound\_ID만을 포함하는 상기 방법과 동일하다. 실시예는 위닝 사운드의 스코어가 다른 모든 사운드 파일의 스코어보다 상당히 높아서 최소 실험으로 적절한 임계값이 선택될 수 있음을 나타낸다. 이 실시예를 실행하는 하나의 방법은 동일한 스코어인

경우 임의의 타이브레이킹으로 최신성에 따라서 데이터베이스 내의 모든 sound\_ID를 분류하는 것이다. 각각의 최신 순위가 고유하기 때문에, 최신 스코어와 sound\_ID 사이에 일 대 일 사상(mapping)이 존재한다. 그 다음, 후보 sound\_ID의 리스트 및 관련된 스캐터 리스트(도 9 참조)를 형성하기 위하여 sound\_ID에 의해 정렬될 때, 상기 순위가 sound\_ID 대신에 사용될 수 있다. 트리플릿(fingerprint, landmark, sound\_ID)의 색인 리스트가 생성될 때 및 상기 색인 리스트가 마스터 색인 리스트로 정렬되기 전에 순위 번호가 색인으로 뛰어오른다. 그 다음, 순위가 sound\_ID의 자리를 차지한다. 대안으로, sound\_ID를 순위로 바꾸기 위해 검색 및 제거 기능이 사용될 수 있다. 순위가 갱신되면, 사상(寫像) 보존이 유지되는 경우 새로운 순위가 이전의 순위로 사상된다.

대안으로, 순위는 프로세스에서 나중에 오를 수 있다. 스캐터 리스트가 생성되면, 순위는 각각의 sound\_ID와 관련될 수 있다. 그 다음, 순위에 의해 세트가 분류된다. 이 실시예에서, 상기 스캐터 리스트로의 포인터들만이 변경될 필요가 있다. 스캐터 리스트로의 그룹화는 반복될 필요가 없다. 추후의 그룹화의 이점은 순위가 갱신될 때마다 전체 데이터베이스 색인이 재생성될 필요가 없다는 것이다.

경제적 가치의 대상으로서 인기 순위 자체가 흥미가 있을 수 있다는 것에 주목하자. 즉, 순위는 알려지지 않은 노래 샘플을 식별하기 위한 소비자들의 바람직함을 반영한다. 많은 경우, 노래 음반을 구매하려는 욕구로 조화가 유발된다. 실제로, 사용자에게 대한 인구 통계학 정보가 알려지면, 각각의 원하는 인구 통계학의 그룹에 대한 다른 순위 결정 방법이 실행될 수 있다. 사용자의 인구 통계학적 그룹은 사용자가 인식 서비스에 서명을 할 때 요구되는 개인 정보로부터 얻어질 수 있다. 또한, 이것은 표준 공동 필터링 기술에 의해 동적으로 결정될 수 있다.

실시간 시스템에서, 사운드는 시간에 대해 증분으로 인식 시스템에 제공되어, 파이프라인식의 인식을 가능하게 한다. 이 경우, 세그먼트로 입력되는 데이터를 처리하여 샘플 색인 세트를 점진적으로 갱신하는 것이 가능하다. 각각의 업데이트 기간 이후, 상기의 검색 및 스캐닝 단계를 사용하여 후보 라이브러리 사운드 레코딩을 검색하기 위하여 새롭게 증가된 색인 세트가 이용된다. 새롭게 얻어진 샘플 핑거프린트와 매칭되는 핑거프린트에 대한 데이터베이스 색인이 검색되고, 새로운 (landmark<sub>k</sub>, landmark<sub>j</sub>, sound\_ID<sub>j</sub>) 트리플릿이 생성된다. 새로운 쌍들이 스캐터 리스트에 추가되어, 막대그래프가 증가된다. 이 방법의 이점은 사운드 레코딩을 명백히 식별하기 위하여 충분한 데이터가 수집되는 경우, 예를 들어 사운드 파일 중의 하나의 오프셋 빈 내의 포인트의 수가 높은 임계값을 초과하거나 그 다음으로 높은 사운드 파일 스코어를 초과하는 경우, 데이터 획득이 종료되고, 그 결과가 알려질 수 있다는 것이다.

정확한 사운드가 식별되었다면, 그 결과가 어떤 적절한 방법으로 사용자 또는 시스템에 보고된다. 예를 들어, 컴퓨터 인쇄 출력, 이메일, 웹 검색 결과 페이지, 이동 전화에 통지되는 SMS(short messaging service) 텍스트, 전화기를 통한 컴퓨터 발생 음성 설명, 또는 사용자가 나중에 액세스할 수 있는 웹 사이트 또는 인터넷 계정으로의 상기 결과의 게시에 의해 상기 결과가 보고될 수 있다. 보고된 결과는 노래의 이름과 가수; 클래식 작품의 작곡가, 곡명 및 레코딩 속성(예를 들어, 연주자, 개최 예정지 등); 광고 회사 및 제품; 또는 다른 적절한 식별자와 같은 사운드의 식별 정보를 포함할 수 있다. 또한, 인명 정보, 주변의 콘서트에 관한 정보, 팬들에 대한 관심 정보가 제공될 수 있다. 이와 같은 데이터에 대한 연결이 제공될 수 있다. 보고된 결과는 또한 사운드 파일의 절대 스코어 또는 그 다음의 가장 높은 스코어를 얻은 파일에 비교되는 그의 스코어를 포함할 수 있다.

인식 방법의 하나의 유용한 결과는 동일한 사운드에 대한 2개의 다른 연주를 혼동하지 않는다는 것이다. 예를 들어, 클래식 음악의 동일한 부분에 대한 다른 공연은 인간이 그 둘 사이의 차이를 검출할 수 없을지라도 동일한 것으로 고려되지 않는다. 이것은 랜드마크/핑거프린트 쌍과 그들의 시간 변화가 2개의 다른 공연에 대해 정확히 매칭되지 않을 것이기 때문이다. 본 실시예에서, 상기 랜드마크/핑거프린트 쌍은 식별될 선행 대응 관계에 대해 서로의 약 10ms 이내에 있어야 한다. 그 결과, 본 발명에 의해 실행되는 자동 인식은 적절한 공연/사운드 트랙 및 예술가/레이블이 모든 경우에 신뢰될 수 있도록 한다.

## 실시예

본 발명의 바람직한 실시예인 연속적인 슬라이딩 윈도우 오디오 인식이 아래에 기술된다. 마이크로폰이나 사운드의 다른 공급원이 지속적으로 버퍼 내로 샘플링되어 이전의 N초의 사운드의 레코딩이 얻어진다. 사운드 버퍼의 콘텐츠는 상기 사운드 콘텐츠의 식별을 확인하기 위하여 주기적으로 분석된다. 사운드 버퍼는 일정한 크기를 가질 수 있거나, 사운드가 샘플링되는 것에 따라서 그 크기가 커질 수 있는데, 본 명세서에서는 오디오 샘플의 순차적으로 성장하는 세그먼트로서 언급된다. 식별된 사운드 레코딩의 존재를 나타내기 위한 보고가 이루어진다. 예를 들어, 로그 파일이 수집될 수 있고, 또는 음악의 곡명, 가수, 앨범 표지, 가사 또는 구매 정보와 같은 음악에 대한 정보를 나타내는 표시 화면이 장치에 나타날 수 있다. 중복(redundancy)을 피하기 위해, 예를 들어 유튜브에서의 프로그램의 변경 이후와 같이 인식된 사운드의 식별이 변경되는 경우에만 보고가 이루어진다. 임의의 사운드 스트림(라디오, 인터넷 스트리밍 라디오, 숨겨진 마이크로폰, 전화 통화 등)으로부터 재생된 음악의 리스트를 생성하기 위하여 이와 같은 장치가 사용될 수 있다. 음악 식별에 추가해서, 인식 시간과 같은 정보가 기록될 수 있다. 위치 정보를 이용할 수 있는 경우(예를 들어, GPS로부터 이용 가능한 경우), 이와 같은 정보가 또한 기록될 수 있다.

식별을 달성하기 위하여, 각각의 버퍼는 새로이(de novo) 식별될 수 있다. 대안으로, 예를 들어 사운드 파라미터가 핑거프린트나 다른 중간 특징 추출 형태로 추출되고, 제 2 버퍼에 저장될 수 있다. 이전의 핑거프린트가 버퍼의 말단으로부터 제거되면서 새로운 핑거프린트가 제 2 버퍼의 전면부에 추가될 수 있다. 이와 같은 회전식의 버퍼 방법의 이점은 사운드 샘플의 이전의 중첩되는 세그먼트에 대해 동일한 분석이 중복되어 실행될 필요가 없으므로 계산상의 노력을 줄일 수 있다는 것이다. 상기 회전식 핑거프린트 버퍼의 콘텐츠에 대한 식별 프로세스가 주기적으로 실행된다. 소형의 휴대용 장치의 경우, 상기 장치에서 핑거프린트 분석이 실행되고, 그 결과가 비교적 낮은 대역폭의 데이터 채널을 사용하여 인식 서버로 전송될 수 있는데, 그 이유는 핑거프린트 스트림이 데이터 집약적이지 않기 때문이다. 회전식의 핑거프린트 버퍼는 휴대용 장치에서 유지될 수 있고, 인식 서버로 매번 이동될 수 있거나, 인식 서버에 유지될 수 있는데, 이 경우 연속적인 인식 세션이 서버에 저장된다.

이와 같은 회전식 버퍼 시스템에서, 인식을 위해 충분한 정보가 이용 가능하게 되면 새로운 사운드 레코딩이 인식될 수 있다. 충분한 정보는 버퍼의 길이 미만의 길이를 가질 수 있다. 예를 들어, 구별이 명확한 노래가 재생 1초 후에 유일하게 인식되고, 시스템이 1초 인식 주기를 갖는 경우, 버퍼가 15-30초 길이를 가질 수 있지만, 상기 노래는 바로 인식될 수 있다. 역으로, 차이가 명확하지 않은 노래가 인식을

위해 보다 많은 시간의 샘플을 필요로 하는 경우, 시스템은 노래의 식별을 선언하기 전에 보다 긴 기간을 기다려야 한다. 이 슬라이딩 윈도우 인식 방법에서, 사운드들은 그들이 식별되자마자 인식될 수 있다.

본 발명이 충분한 기능의 인식 시스템과 방법의 문맥으로 설명되었지만, 당업자들은 본 발명의 메커니즘이 다양한 형태의 명령들의 컴퓨터 기록 가능한 매체의 형태로 분배될 수 있고, 본 발명이 상기 분배를 실제로 실행하기 위해 사용된 특정한 타입의 신호 베어링 매체에 관계 없이 동일하게 적용됨을 이해할 수 있을 것이다. 이와 같은 컴퓨터 액세스 가능한 장치의 실례로는 컴퓨터 메모리(RAM 또는 ROM), 플로피 디스크, CD-ROM뿐만 아니라, 디지털과 아날로그 통신 링크와 같은 전송 타입의 매체 등이 있다.

#### (57) 청구의 범위

##### 청구항 1.

매체 샘플과 매체 파일을 비교하는 방법에 있어서,  
각각이 상기 매체 샘플 내의 특정한 위치의 특성을 나타내는 한 세트의 샘플 핑거프린트(fingerprint)를 결정하는 단계,  
각각이 상기 매체 파일 내의 적어도 하나의 파일 위치의 특징을 나타내는 한 세트의 파일 핑거프린트를 획득하는 단계,  
상기 매체 샘플의 상기 특정 위치와 상기 매체 파일의 상기 파일 위치 사이의 대응 관계를 생성하는 단계 - 여기서, 대응하는 위치들은 동일한 핑거프린트를 가짐, 및  
복수의 상기 대응하는 위치들이 실질적으로 선형적으로 관련되는 경우, 상기 매체 샘플과 상기 매체 파일을 식별하는 단계를 포함하는 것을 특징으로 하는 매체 샘플과 매체 파일의 비교 방법.

##### 청구항 2.

오디오 샘플과 오디오 파일을 비교하는 방법에 있어서,  
적어도 하나의 오디오 파일의 각각에 대하여, 상기 오디오 파일을 나타내는 복수의 파일 핑거프린트를 계산하는 단계,  
상기 오디오 샘플을 나타내는 복수의 샘플 핑거프린트를 계산하는 단계, 및  
적어도 상기 파일 핑거프린트의 임계 번호가 상기 샘플 핑거프린트와 같은 경우, 상기 오디오 샘플과 상기 오디오 파일을 식별하는 단계를 포함하고,  
상기 샘플 핑거프린트는 상기 오디오 샘플의 시간 연장에 대해 불변인 것을 특징으로 하는 오디오 샘플과 오디오 파일의 비교 방법.

##### 청구항 3.

오디오 샘플의 특징을 나타내는 방법에 있어서,  
상기 오디오 샘플 내의 한 세트의 재생 가능한 위치들을 계산하는 단계, 및  
상기 오디오 샘플 내의 상기 재생 가능한 위치들의 특징을 나타내는 한 세트의 핑거프린트를 계산하는 단계를 포함하는 것을 특징으로 하는 오디오 샘플의 특징을 나타내는 방법.

##### 청구항 4.

오디오 샘플의 특징을 나타내는 방법에 있어서,  
상기 오디오 샘플의 스펙트로그램으로부터 적어도 하나의 핑거프린트를 계산하는 단계를 포함하고,  
상기 스펙트로그램은 앵커 돌출점(anchor salient point)과 연결된 돌출점을 포함하고, 상기 핑거프린트는 상기 앵커 돌출점과 임의의 연결된 돌출점의 주파수 좌표로부터 계산되는 것을 특징으로 하는 오디오 샘플의 특징을 나타내는 방법.

##### 청구항 5.

매체 샘플을 인식하는 방법에 있어서,  
상기 매체 파일과 상기 매체 샘플의 실질적인 복수의 동등한 특성의 위치가 실질적으로 선형으로 관련되는 매체 파일들을 식별하는 단계를 포함하는 것을 특징으로 하는 매체 샘플 인식 방법.

##### 청구항 6.

매체 샘플을 인식하는 방법에 있어서,  
복수의 매체 파일의 각각에 대하여, 상기 매체 파일의 파일 표현을 제공하는 단계,  
상기 매체 샘플의 샘플 표현을 제공하는 단계, 및  
상기 파일 표현들 중에서 적어도 하나의 동일한 파일 표현을 식별하는 단계를 포함하고,  
상기 파일 표현을 검색함으로써 상기 동일한 파일 표현이 상기 샘플 표현과 동일하고, 상기 검색은 상기 파일 표현의 식별의 확률에 따라서 부분적으로 실행되는 것을 특징으로 하는 매체 샘플 인식 방법.

##### 청구항 7.

매체 샘플을 인식하는 방법에 있어서,  
상기 매체 샘플의 세그먼트의 특징을 나타내는 한 세트의 샘플 핑거프린트를 계산하는 단계,  
상기 핑거프린트를 회전식 버퍼에 저장하는 단계,  
각각이 적어도 하나의 매체 파일의 특징을 나타내고 상기 회전식 버퍼 내의 적어도 하나의 핑거프린트와 매칭되는 한 세트의 매칭 핑거프린트를 데이터베이스 색인 내에서 획득하는 단계,

복수의 매칭 핑거프린트를 갖는 적어도 하나의 매체 파일을 식별하는 단계, 및  
 상기 회전식 버퍼로부터 적어도 하나의 샘플 핑거프린트를 제거하는 단계를 포함하는 것을 특징으로 하는 매체 샘플 인식 방법.

#### 청구항 8.

데이터베이스 내에 적어도 하나의 오디오 파일의 데이터베이스 색인을 생성하는 컴퓨터 실행 방법에 있어서,  
 각각이 상기 오디오 파일 내의 특정한 위치의 특성을 나타내고, 각각의 오디오 파일의 특징을 표현하는 한 세트의 핑거프린트를 계산하는 단계, 및

메모리 내에 상기 핑거프린트들, 상기 위치들 및 각각의 매체 파일의 식별자를 저장하는 단계를 포함하고,  
 각각의 대응하는 핑거프린트, 위치 및 식별자는 상기 메모리 내에서 관련되는 것을 특징으로 하는 컴퓨터 실행 방법.

#### 청구항 9.

제 1 항에 있어서,  
 상기 결정 단계는 한 세트의 샘플 핑거프린트를 계산하는 단계를 포함하는 것을 특징으로 하는 매체 샘플과 매체 파일의 비교 방법.

#### 청구항 10.

제 1 항에 있어서,  
 상기 결정 단계는 한 세트의 샘플 핑거프린트를 수신하는 단계를 포함하는 것을 특징으로 하는 매체 샘플과 매체 파일의 비교 방법.

#### 청구항 11.

오디오 샘플과 오디오 파일을 비교하는 방법에 있어서,  
 적어도 하나의 오디오 파일의 각각에 대하여, 상기 오디오 파일을 나타내는 복수의 핑거프린트를 계산하는 단계,  
 상기 오디오 샘플을 나타내는 복수의 샘플 핑거프린트를 계산하는 단계, 및  
 적어도 상기 파일 핑거프린트의 임계 번호가 상기 샘플 핑거프린트와 동일한 경우 상기 오디오 샘플과 상기 오디오 파일을 식별하는 단계를 포함하고,  
 각각의 샘플 핑거프린트는 청구항 4의 방법에 따라서 상기 오디오 샘플의 스펙트로그램으로부터 계산되는 것을 특징으로 하는 오디오 샘플과 오디오 파일의 비교 방법.

#### 청구항 12.

제 9 항에 있어서,  
 상기 매체 샘플은 오디오 샘플인 것을 특징으로 하는 매체 샘플과 매체 파일의 비교 방법.

#### 청구항 13.

제 9 항에 있어서,  
 상기 식별 단계는 상기 대응하는 위치들의 산점도 내에서 대각선을 찾는 단계를 포함하는 것을 특징으로 하는 매체 샘플과 매체 파일의 비교 방법.

#### 청구항 14.

제 13 항에 있어서,  
 상기 대각선을 찾는 단계는 상기 대응하는 위치들 사이의 차이를 형성하는 단계를 포함하는 것을 특징으로 하는 매체 샘플과 매체 파일의 비교 방법.

#### 청구항 15.

제 14 항에 있어서,  
 상기 대각선을 찾는 단계는 상기 차이를 정렬하는 단계를 추가로 포함하는 것을 특징으로 하는 매체 샘플과 매체 파일의 비교 방법.

#### 청구항 16.

제 14 항에 있어서,  
 상기 대각선을 찾는 단계는 상기 차이의 막대그래프의 피크를 계산하는 단계를 추가로 포함하는 것을 특징으로 하는 매체 샘플과 매체 파일의 비교 방법.

#### 청구항 17.

제 9 항에 있어서,  
 상기 식별 단계는 상기 대응 관계의 허프 변환(Hough transform)과 라돈 변환(Radon transform) 중의 하나를 계산하는 단계를 포함하는 것을 특징으로 하는 매체 샘플과 매체 파일의 비교 방법.

#### 청구항 18.

제 17 항에 있어서,  
 상기 식별 단계는 상기 허프 또는 라돈 변환의 피크를 찾는 단계를 추가로 포함하는 것을 특징으로 하는 매체 샘플과 매체 파일의 비교

방법.

#### 청구항 19.

제 9 항에 있어서,

상기 식별 단계는 상기 대응 관계의 수가 임계값을 초과하는지를 결정하는 단계를 포함하는 것을 특징으로 하는 매체 샘플과 매체 파일의 비교 방법.

#### 청구항 20.

제 9 항에 있어서,

데이터베이스 색인으로부터 추가적인 매체 파일의 파일 위치의 특징을 나타내는 추가적인 핑거프린트를 획득하는 단계,

상기 매체 샘플의 상기 특정 위치와 상기 추가적인 매체 파일의 상기 파일 위치 사이의 추가적인 대응 관계를 생성하는 단계 - 여기서, 대응하는 위치는 동일한 핑거프린트를 가짐, 및

상당히 많은 복수의 사실상 선형으로 관련되는 대응 위치를 갖는 위닝 매체 파일(winning media file)을 선택하는 단계를 포함하는 것을 특징으로 하는 매체 샘플과 매체 파일의 비교 방법.

#### 청구항 21.

제 20 항에 있어서,

복수의 상기 대응 위치가 사실상 선형으로 관련되는 매체 파일을 식별하는 단계를 추가로 포함하고, 상기 선택 단계는 상기 식별된 매체 파일로부터 위닝 매체 파일을 선택하는 단계를 포함하는 것을 특징으로 하는 매체 샘플과 매체 파일의 비교 방법.

#### 청구항 22.

제 21 항에 있어서,

상기 복수의 대응 위치가 사실상 선형으로 관련되는 매체 파일을 식별하는 단계는 상기 추가적인 매체 파일의 제 1 서브세트를 검색하는 단계를 포함하는 것을 특징으로 하는 매체 샘플과 매체 파일의 비교 방법.

#### 청구항 23.

제 22 항에 있어서,

상기 제 1 서브세트 내의 추가 매체 파일은 상기 제 1 서브세트 내에 있지 않은 추가 매체 파일보다 식별될 확률이 높은 것을 특징으로 하는 매체 샘플과 매체 파일의 비교 방법.

#### 청구항 24.

제 22 항에 있어서,

상기 복수의 대응 위치가 사실상 선형으로 관련되는 매체 파일을 식별하는 단계는 상기 추가 매체 파일의 제 2 서브세트를 검색하는 단계를 추가로 포함하고, 상기 제 2 서브세트는 상기 제 1 서브세트 내의 매체 파일이 식별되지 않는 경우에 검색되는 것을 특징으로 하는 매체 샘플과 매체 파일의 비교 방법.

#### 청구항 25.

제 21 항에 있어서,

식별될 확률에 따라서 상기 추가 매체 파일을 배열하는 단계를 추가로 포함하는 것을 특징으로 하는 매체 샘플과 매체 파일의 비교 방법.

#### 청구항 26.

제 25 항에 있어서,

상기 복수의 대응하는 위치가 사실상 선형으로 관련되는 매체 파일을 식별하는 단계는 상기 배열에 따라서 상기 추가 매체 파일을 검색하는 단계를 포함하는 것을 특징으로 하는 매체 샘플과 매체 파일의 비교 방법.

#### 청구항 27.

제 21 항에 있어서,

상기 복수의 대응하는 위치가 사실상 선형으로 관련되는 매체 파일을 식별하는 단계는 소정의 임계값을 초과하는, 상기 복수의 사실상 선형으로 관련되는 대응 위치를 갖는 매체 파일에서 상기 검색을 종료하는 단계를 포함하는 것을 특징으로 하는 매체 샘플과 매체 파일의 비교 방법.

#### 청구항 28.

제 9 항에 있어서,

상기 방법은 분산 시스템에서 실행되는 것을 특징으로 매체 샘플과 매체 파일의 비교 방법.

#### 청구항 29.

제 28 항에 있어서,

상기 계산 단계는 클라이언트 장치에서 실행되고, 상기 획득 단계, 생성 단계 및 식별 단계는 중앙 위치에서 실행되며, 상기 방법은 상기

샘플 핑거프린트를 상기 클라이언트 장치로부터 상기 중앙 위치로 전송하는 단계를 추가로 포함하는 것을 특징으로 하는 매체 샘플과 매체 파일의 비교 방법.

**청구항 30.**

제 9 항에 있어서,

순차적으로 성장하는 상기 매체 샘플의 세그먼트에 대한 상기 계산 단계, 획득 단계, 생성 단계 및 식별 단계를 반복하는 단계를 추가로 포함하는 것을 특징으로 하는 매체 샘플과 매체 파일의 비교 방법.

**청구항 31.**

제 9 항에 있어서,

상기 획득 단계, 생성 단계 및 식별 단계는 상기 계산된 핑거프린트를 저장하는 회전식 버퍼에서 주기적인 간격으로 실행되는 것을 특징으로 하는 매체 샘플과 매체 파일의 비교 방법.

**청구항 32.**

제 9 항에 있어서,

상기 매체 샘플을 획득하는 단계를 추가로 포함하고, 상기 계산 단계와 상기 획득 단계는 동시에 실행되는 것을 특징으로 하는 매체 샘플과 매체 파일의 비교 방법.

**청구항 33.**

제 8 항에 있어서,

핑거프린트 값에 의해 상기 데이터베이스 색인을 정렬하는 단계를 추가로 포함하는 것을 특징으로 하는 컴퓨터 실행 방법.

**청구항 34.**

제 8 항 또는 제 9 항에 있어서,

상기 각각의 오디오 파일 또는 상기 매체 샘플 내의 특정 위치는 상기 오디오 파일 또는 매체 샘플에 따라서 계산되는 것을 특징으로 하는 방법.

**청구항 35.**

제 8 항 또는 제 9 항에 있어서,

각각의 핑거프린트는 상기 특정 위치 부근의 상기 오디오 파일 또는 매체 샘플의 적어도 하나의 특징을 나타내는 것을 특징으로 하는 방법.

**청구항 36.**

제 8 항 또는 제 9 항에 있어서,

상기 핑거프린트는 숫자로 표시되는 값인 것을 특징으로 하는 방법.

**청구항 37.**

제 8 항 또는 제 12 항에 있어서,

상기 핑거프린트의 값은 상기 핑거프린트를 계산하기 위한 방법을 지정하는 것을 특징으로 하는 방법.

**청구항 38.**

제 8 항 또는 제 12 항에 있어서,

상기 특정 위치들은 상기 오디오 파일 또는 샘플 내의 타임포인트인 것을 특징으로 하는 방법.

**청구항 39.**

제 38 항에 있어서,

상기 타임포인트는 상기 오디오 파일 또는 샘플의 스펙트럼  $L_p$  놈(norm)의 극대에서 발생하는 것을 특징으로 하는 방법.

**청구항 40.**

제 8 항 또는 제 12 항에 있어서,

상기 핑거프린트는 상기 오디오 파일 또는 샘플의 주파수 분석으로부터 계산되는 것을 특징으로 하는 방법.

**청구항 41.**

제 8 항 또는 제 12 항에 있어서,

상기 핑거프린트는 스펙트럼 슬라이스 핑거프린트, LPC 계수, 셉스트럴 계수(cepstral coefficient)로 구성된 그룹으로부터 선택되는 것을 특징으로 하는 방법.

**청구항 42.**

제 8 항 또는 제 12 항에 있어서,



상기 핑거프린트는 상기 오디오 파일 또는 샘플의 스펙트로그램으로부터 계산되는 것을 특징으로 하는 방법.

#### 청구항 43.

제 42 항에 있어서,

상기 스펙트로그램의 돌출점은 시간 좌표와 주파수 좌표를 포함하고, 상기 특정 위치들은 상기 시간 좌표로부터 계산되고, 상기 핑거프린트들은 상기 주파수 좌표로부터 계산되는 것을 특징으로 하는 방법.

#### 청구항 44.

제 43 항에 있어서,

상기 복수의 돌출점을 앵커 돌출점에 연결하는 단계를 추가로 포함하고, 상기 특정 위치들 중의 하나는 상기 앵커 돌출점의 시간 좌표로부터 계산되고, 대응하는 핑거프린트는 상기 연결된 돌출점과 상기 앵커 포인트 중의 적어도 하나의 주파수 좌표로부터 계산되는 것을 특징으로 하는 방법.

#### 청구항 45.

제 44 항에 있어서,

상기 대응하는 핑거프린트는 상기 연결된 돌출점과 상기 앵커 포인트의 2개의 상기 주파수 좌표 사이의 몫(quotient)으로부터 계산되고, 상기 대응하는 핑거프린트는 시간 연장에 대해 불변인 것을 특징으로 하는 방법.

#### 청구항 46.

제 45 항에 있어서,

상기 대응하는 핑거프린트는 상기 앵커 포인트의 상기 시간 좌표와 상기 연결된 돌출점의 상기 시간 좌표 사이의 적어도 하나의 시간 차이로부터 추가로 계산되는 것을 특징으로 하는 방법.

#### 청구항 47.

제 46 항에 있어서,

상기 대응하는 핑거프린트는 상기 연결된 돌출점 및 상기 앵커 포인트의 상기 주파수 좌표 중 하나와 상기 시간 차이 중 하나의 곱으로부터 계산되고, 상기 대응하는 핑거프린트는 시간 연장에 대해 불변인 것을 특징으로 하는 방법.

#### 청구항 48.

제 8 항 또는 제 12 항에 있어서,

상기 특정 위치들과 상기 핑거프린트들은 상기 오디오 파일 또는 샘플의 다중 차원의 함수의 돌출점으로부터 계산되고, 상기 차원 중의 적어도 하나는 시간 차원이고, 상기 차원 중의 적어도 하나는 비시간 차원인 것을 특징으로 하는 방법.

#### 청구항 49.

제 48 항에 있어서,

상기 특정 위치들은 상기 시간 차원으로부터 계산되는 것을 특징으로 하는 방법.

#### 청구항 50.

제 48 항에 있어서,

상기 핑거프린트들은 상기 비시간 차원 중의 적어도 하나로부터 계산되는 것을 특징으로 하는 방법.

#### 청구항 51.

제 11 항 또는 제 48 항에 있어서,

상기 돌출점은 상기 다중 차원 함수의 극대, 극소 및 제로 교차점으로 구성된 그룹으로부터 선택되는 것을 특징으로 하는 방법.

#### 청구항 52.

제 8 항 또는 제 12 항에 있어서,

상기 핑거프린트들은 시간 연장에 대해 불변인 것을 특징으로 하는 방법.

#### 청구항 53.

제 8 항 또는 제 12 항에 있어서,

각각의 핑거프린트는 상기 오디오 파일 또는 샘플의 다수의 타임슬라이스로부터 계산되는 것을 특징으로 하는 방법.

#### 청구항 54.

제 53 항에 있어서,

상기 다수의 타임슬라이스는 가변의 시간 양에 의해 오프셋되는 것을 특징으로 하는 방법.

#### 청구항 55.

제 54 항에 있어서,

상기 핑거프린트들은 상기 가변의 양으로부터 부분적으로 계산되는 것을 특징으로 하는 방법.

**청구항 56.**

제 6 항에 있어서,

상기 적어도 하나의 동일한 파일 표현은 상기 샘플 표현에 대한 임계값 유사성을 초과하는 것을 특징으로 하는 방법.

**청구항 57.**

제 6 항에 있어서,

상기 식별 단계는 상기 파일 표현의 제 1 서브세트를 검색하는 단계를 포함하고, 상기 제 1 서브세트는 상기 제 1 서브세트 내에 있지 않은 파일 표현보다 높은 식별 확률을 갖는 파일 표현을 포함하는 것을 특징으로 하는 방법.

**청구항 58.**

제 57 항에 있어서,

상기 제 1 서브세트가 상기 적어도 하나의 동일한 파일 표현을 포함하는 않는 경우 상기 파일 표현의 제 2 서브세트를 검색하는 단계를 추가로 포함하는 것을 특징으로 하는 방법.

**청구항 59.**

제 6 항에 있어서,

상기 식별 확률에 의해 상기 파일 표현을 배열하는 단계를 추가로 포함하고, 상기 식별 단계는 상기 배열 순서로 상기 파일 표현을 검색하는 단계를 포함하는 것을 특징으로 하는 매체 샘플 인식 방법.

**청구항 60.**

제 59 항에 있어서,

상기 적어도 하나의 동일한 파일 표현이 식별되는 경우 상기 검색을 종료하는 단계를 추가로 포함하는 것을 특징으로 하는 매체 샘플 인식 방법.

**청구항 61.**

제 6 항, 제 23 항 또는 25 항 중 어느 한 항에 있어서,

상기 식별 확률은 이전의 식별의 최신성(recency)에 따라서 부분적으로 계산되는 것을 특징으로 하는 방법.

**청구항 62.**

제 61 항에 있어서,

특정 파일 표현의 최신 스코어는 상기 특정 파일 표현이 식별될 때 증가되는 것을 특징으로 하는 방법.

**청구항 63.**

제 61 항에 있어서,

상기 파일 표현의 최신 스코어는 일정한 시간 간격으로 감소되는 것을 특징으로 하는 방법.

**청구항 64.**

제 63 항에 있어서,

상기 최신 스코어는 시간에 따라 지수함수적으로 감소되는 것을 특징으로 하는 방법.

**청구항 65.**

제 6 항 또는 제 23 항에 있어서,

상기 식별 확률은 이전의 인식의 주파수에 따라서 부분적으로 계산되는 것을 특징으로 하는 방법.

**청구항 66.**

제 2 항에 있어서,

상기 샘플 핑거프린트는 상기 오디오 샘플의 주파수 성분의 뒀을 포함하는 것을 특징으로 하는 오디오 샘플과 오디오 파일을 비교하는 방법.

**청구항 67.**

제 2 항에 있어서,

상기 샘플 핑거프린트는 상기 오디오 샘플의 주파수 성분과 상기 오디오 샘플의 포인트들 사이의 시간 차의 곱을 포함하는 것을 특징으로 하는 방법.

**청구항 68.**

제 4 항, 제 11 항 또는 제 44 항 중 어느 한 항에 있어서,

상기 연결된 돌출 점들은 목표 지역 내로 내려가는 것을 특징으로 방법.

**청구항 69.**

제 68 항에 있어서,  
상기 목표 지역은 시간 범위에 의해 정의되는 것을 특징으로 하는 방법.

**청구항 70.**

제 68 항에 있어서,  
상기 목표 지역은 주파수 범위에 의해 정의되는 것을 특징으로 하는 방법.

**청구항 71.**

제 68 항에 있어서,  
상기 목표 지역은 변경 가능한 것을 특징으로 하는 방법.

**청구항 72.**

제 7 항에 있어서,  
상기 매체 샘플의 추가 세그먼트에 대하여 상기 방법을 반복하는 단계를 추가로 포함하는 것을 특징으로 하는 매체 샘플 인식 방법.

**청구항 73.**

제 7 항에 있어서,  
상기 계산 단계, 저장 단계 및 제거 단계는 클라이언트 장치에서 실행되고, 상기 획득 단계 및 식별 단계는 중앙 위치에서 실행되며, 상기 방법은 상기 클라이언트 장치로부터 상기 중앙 위치로 상기 샘플 핑거프린트를 전송하는 단계를 추가로 포함하는 것을 특징으로 하는 매체 샘플 인식 방법.

**청구항 74.**

제 7 항에 있어서,  
상기 계산 단계는 클라이언트 장치에서 실행되고, 상기 저장 단계, 획득 단계, 식별 단계 및 제거 단계는 중앙 위치에서 실행되며, 상기 방법은 상기 클라이언트 장치로부터 상기 중앙 위치로 상기 핑거프린트를 전송하는 단계를 추가로 포함하는 것을 특징으로 하는 매체 샘플 인식 방법.

**청구항 75.**

제 3 항에 있어서,  
상기 재생 가능한 위치와 상기 핑거프린트는 동시에 계산되는 것을 특징으로 하는 방법.

**청구항 76.**

컴퓨터에 의해 액세스 가능하고, 매체 샘플과 매체 파일을 비교하기 위한 방법의 단계를 실행하기 위하여 상기 컴퓨터에 의해 실행될 수 있는 명령 프로그램을 실제적으로 구현하는 프로그램 기억 장치에 있어서,

상기 방법의 단계는,

각각이 상기 매체 샘플 내의 특정한 위치의 특징을 나타내는 한 세트의 샘플 핑거프린트를 계산하는 단계,

각각이 상기 매체 파일 내의 적어도 하나의 파일 위치의 특징을 나타내는 한 세트의 핑거프린트를 획득하는 단계,

상기 매체 샘플의 상기 특정 위치와 상기 매체 파일의 상기 파일 위치 사이의 대응관계를 생성하는 단계 - 여기서, 대응하는 위치는 동일한 핑거프린트를 가짐, 및

복수의 상기 대응하는 위치가 사실상 선형으로 관련되는 경우, 상기 매체 샘플과 상기 매체 파일을 식별하는 단계를 포함하는 것을 특징으로 하는 프로그램 기억 장치.

**청구항 77.**

매체 샘플을 인식하기 위한 시스템에 있어서,

상기 매체 샘플 내의 한 세트의 특정 위치와 각각이 상기 특정 위치 중의 하나의 특징을 나타내는 한 세트의 샘플 핑거프린트를 계산하기 위한 랜드마킹 및 핑거프린팅 대상,

적어도 하나의 매체 파일에 대해 대응하는 파일 핑거프린트와 파일 위치를 포함하는 데이터베이스 색인, 및

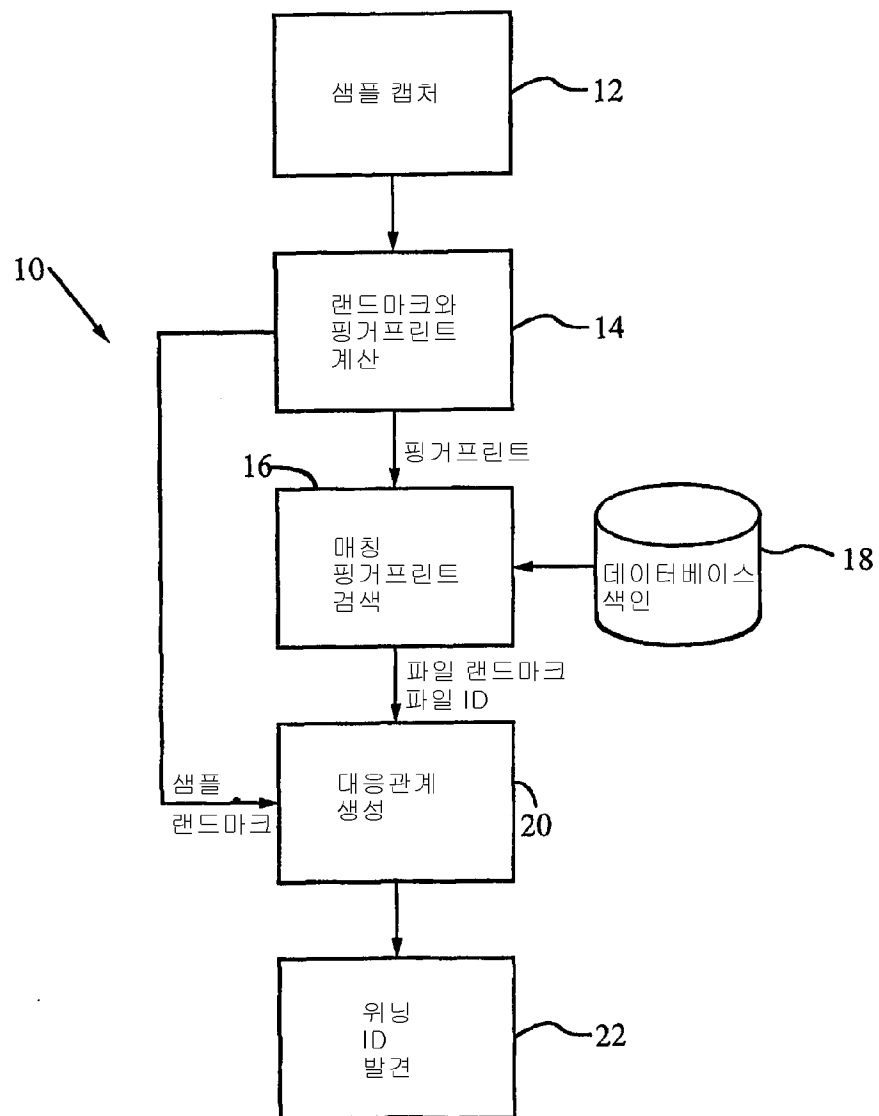
상기 샘플 핑거프린트와 동일한, 상기 데이터베이스 색인 내의 한 세트의 매칭 핑거프린트를 찾고,

상기 매체 샘플의 상기 특정 위치와 상기 적어도 하나의 매체 파일의 파일 위치 사이의 대응 관계 - 여기서, 대응하는 위치는 동일한 핑거프린트를 가짐 - 을 생성하며,

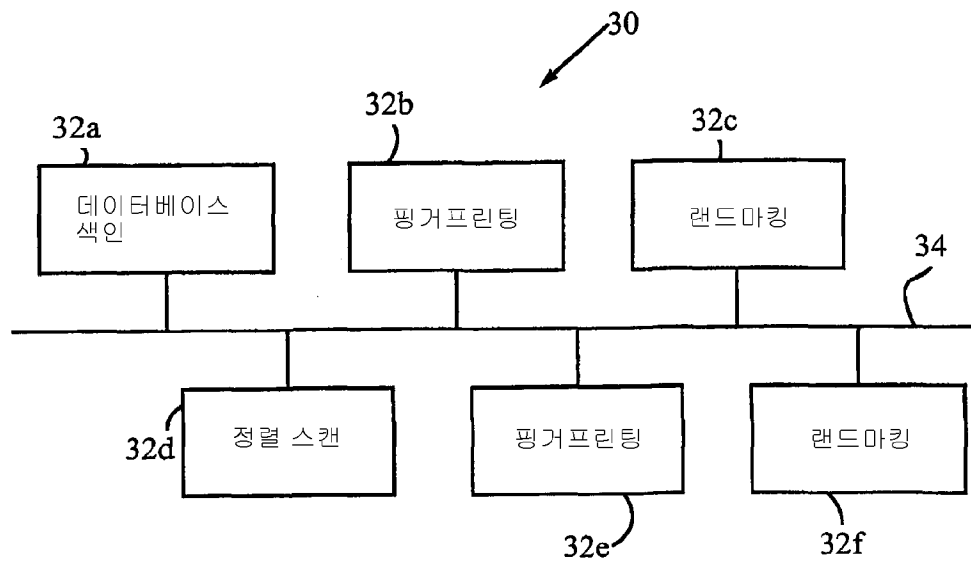
복수의 상기 대응하는 위치가 사실상 선형적으로 관련되는 적어도 하나의 매체 파일을 식별하기 위한 분석 객체를 포함하는 것을 특징으로 하는 매체 샘플을 인식하기 위한 시스템.

도면

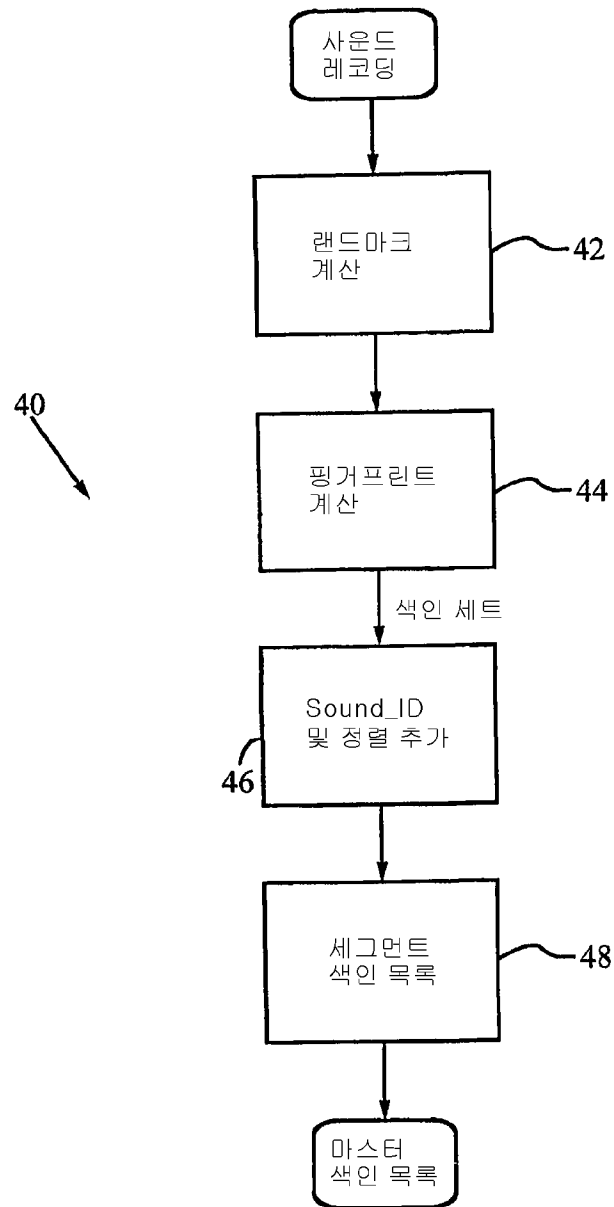
도면 1



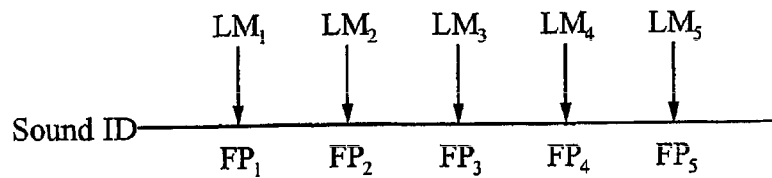
도면 2



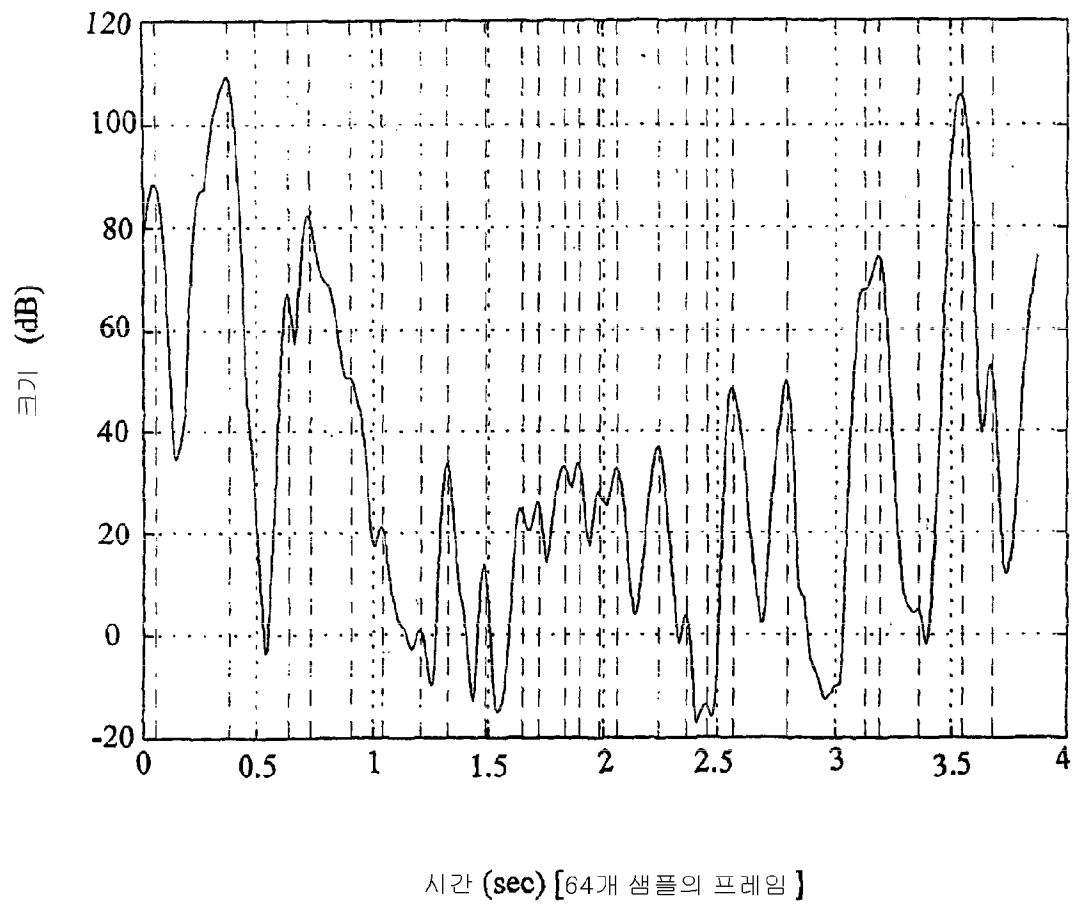
도면 3



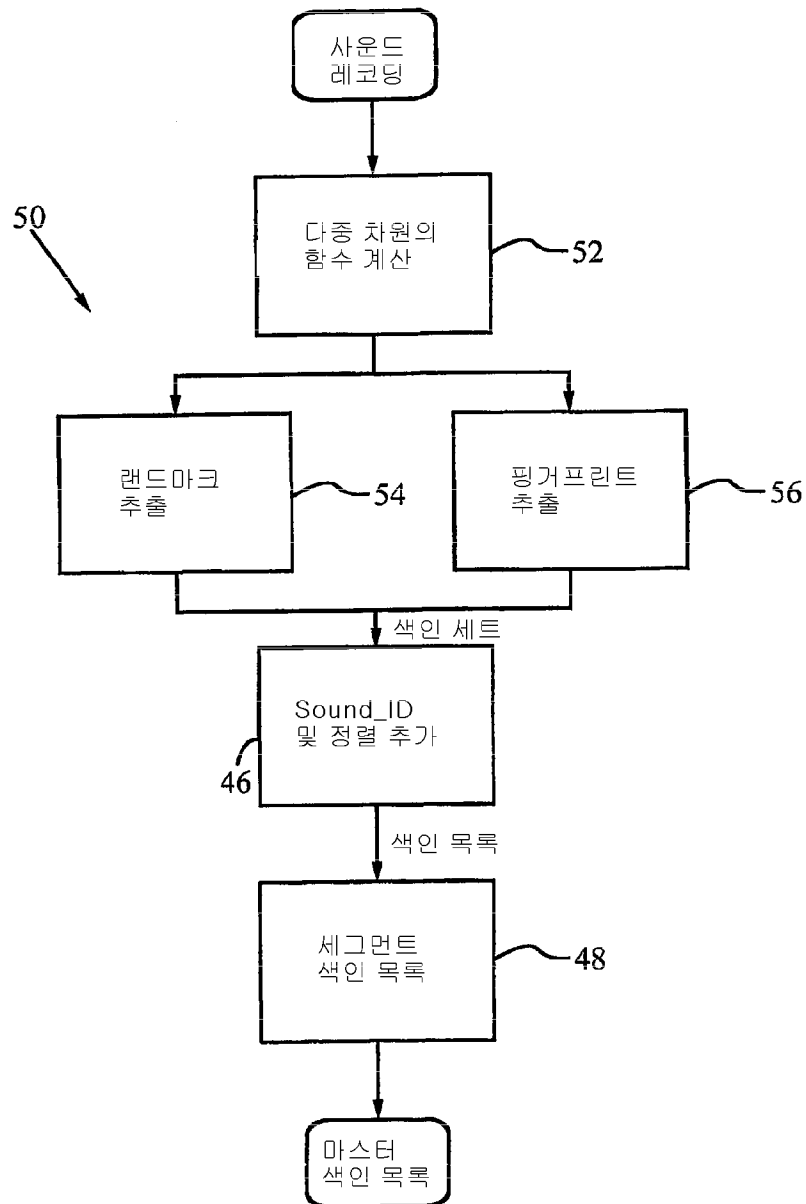
도면 4



도면 5

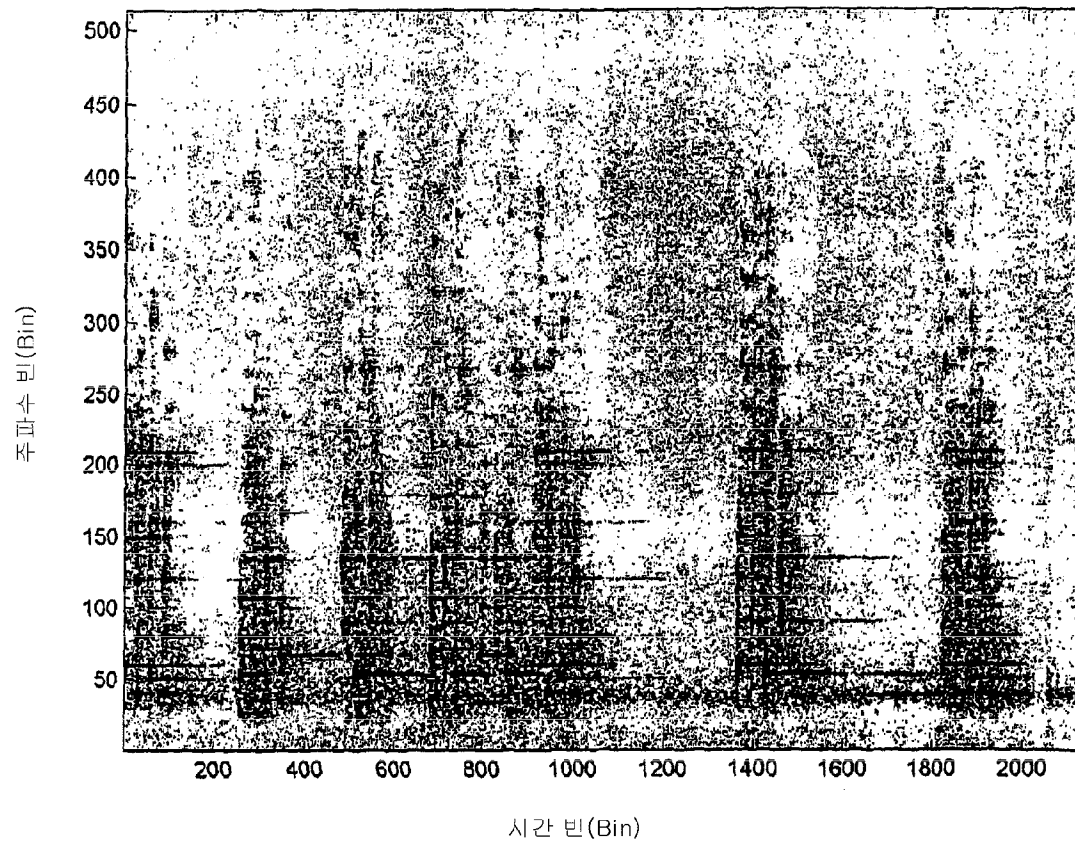


도면 6

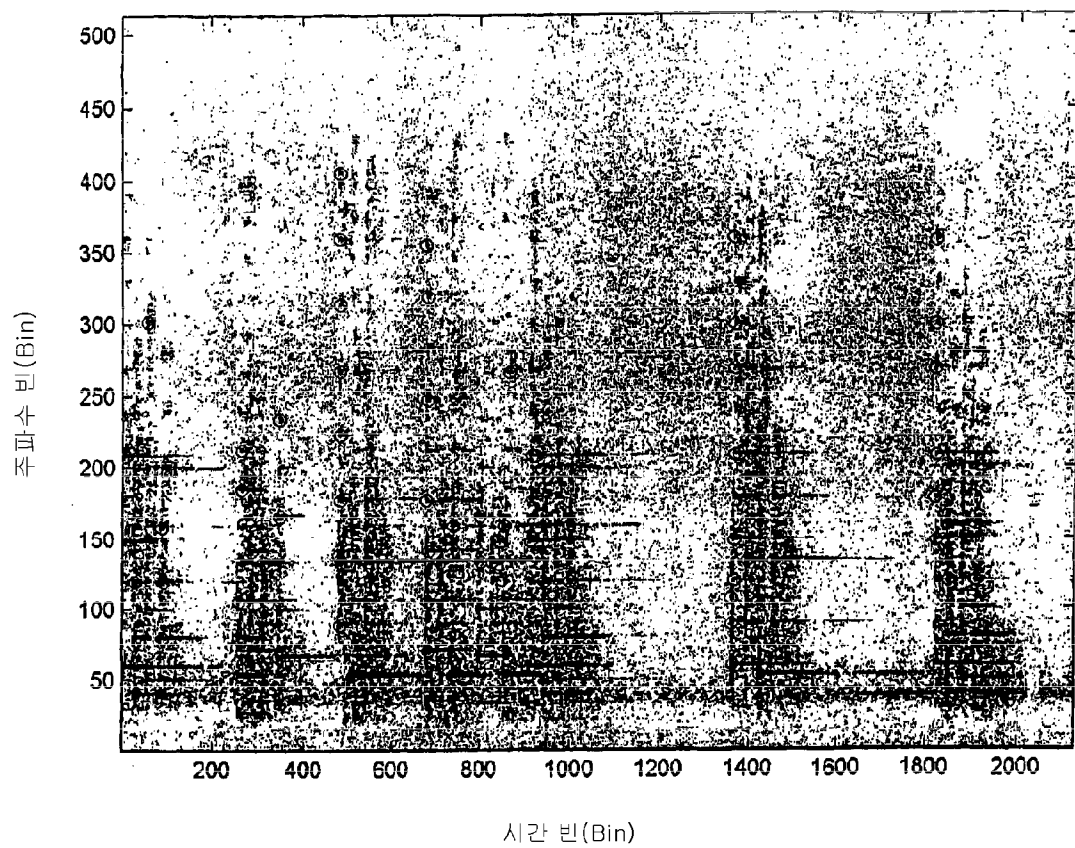




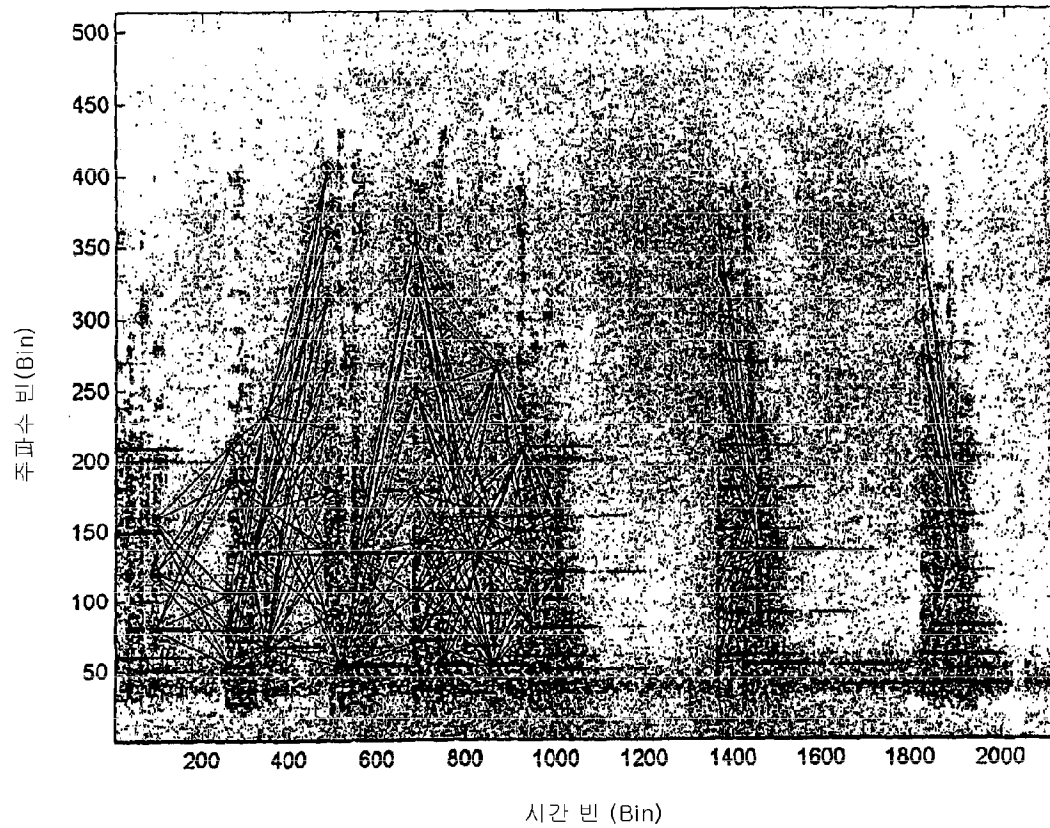
도면 7A



도면 7B



도면 7C



도면 8A

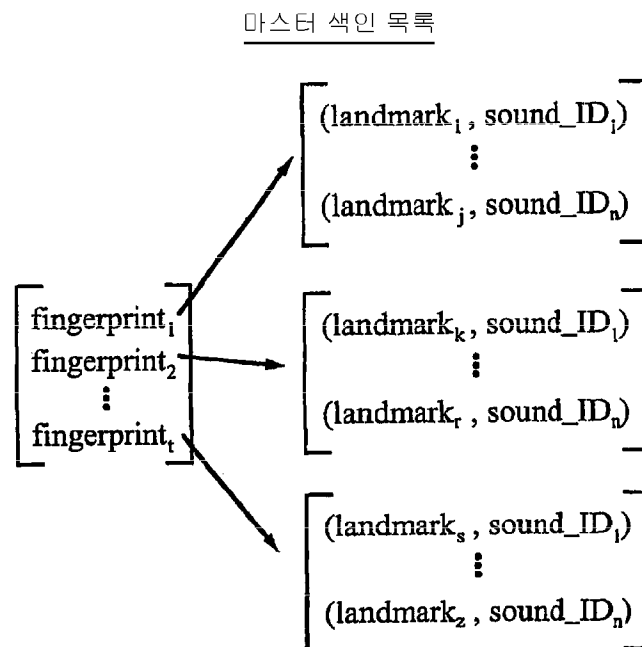
### 색인 세트

$$\begin{array}{ccc}
 \text{Sound\_ID}_1 & & \text{Sound\_ID}_n \\
 \left[ \begin{array}{c} (\text{fingerprint}_1, \text{landmark}_1) \\ (\text{fingerprint}_2, \text{landmark}_2) \\ \vdots \\ (\text{fingerprint}_m, \text{landmark}_m) \end{array} \right] & \dots & \left[ \begin{array}{c} (\text{fingerprint}_p, \text{landmark}_p) \\ (\text{fingerprint}_{p+1}, \text{landmark}_{p+1}) \\ \vdots \\ (\text{fingerprint}_t, \text{landmark}_t) \end{array} \right]
 \end{array}$$

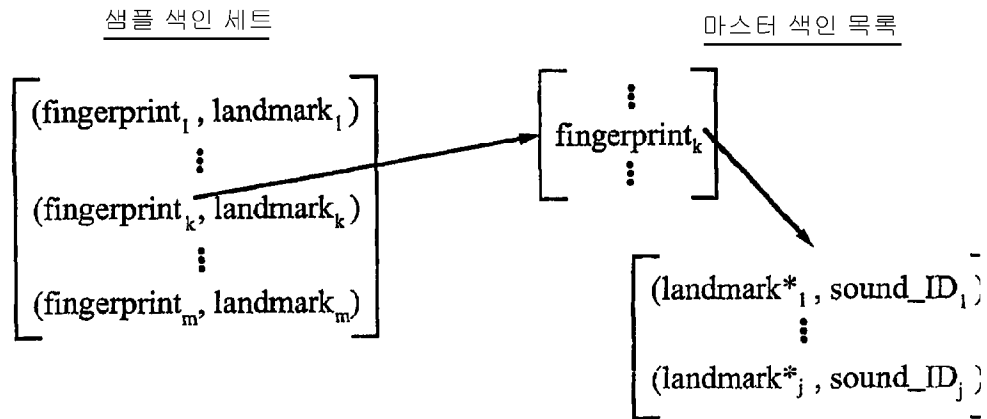
도면 8B

**FIG. 8B**

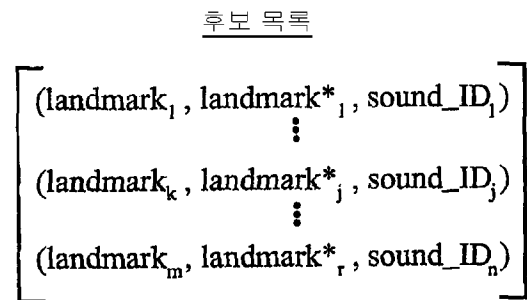
도면 8C



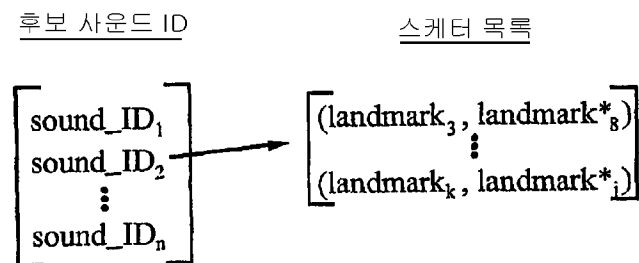
도면 9A



도면 9B



도면 9C



사운드 샘플 시간

사운드 파일 시간

도면 10B

발견되지 않은 위닝 파일: 발견되지 않은 선형 대응

